

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2001, 15:50:48 ; Search time 24.95 Seconds  
(without alignments)  
1061.683 Million cell updates/sec

Title: US-08-978-174-1  
Perfect score: 1185  
Sequence: 1 MGPLRTVELFDVLSFYSW.....AHLGKWMGPIPPAVNARL 226

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL.15.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1185	100.0	226	4 Q9Y2Q3	Q9Y2Q3 homo sapien
2	1169	98.6	226	4 Q9P1S4	Q9P1S4 homo sapien
3	164.5	13.9	194	2 Q52782	Q52782 rhizobium 1
4	149	12.6	195	2 Q9X9Q7	Q9X9Q7 spingomona
5	145.5	12.3	312	5 Q22312	Q22312 caenorhabdi
6	140	11.8	197	2 Q85994	Q85994 spingomona
7	135	11.4	196	2 Q9ZHH4	Q9ZHH4 burkholderi
8	114	9.6	196	2 Q9WXH0	Q9WXH0 alcaligenes
9	106	8.9	199	2 Q9Z3X5	Q9Z3X5 pseudomonas
10	97.5	8.2	238	2 Q9Z167	Q9Z167 pseudomonas
11	91	7.7	581	2 Q9WXQ0	Q9WXQ0 thermocoga
12	90.5	7.6	410	1 Q26322	Q26322 methanobact
13	89	7.5	199	2 Q51499	Q51499 pseudomonas
14	89	7.5	592	10 Q9SPF1	Q9SPF1 arabidopsis
15	89	7.5	592	10 Q9ZSS6	Q9ZSS6 arabidopsis
16	83.5	7.0	308	2 Q9ZNM7	Q9ZNM7 exiguobacte
17	83.5	7.0	1007	10 Q9ZVD4	Q9ZVD4 arabidopsis
18	83	7.0	863	12 Q93124	Q93124 human calic
19	82.5	7.0	263	2 Q87197	Q87197 thermus aqu

007298 pseudomonas	203	6.9	82	20	007298
052799 rhizobium 1	296	6.9	82	21	052799
055098 mus musculus	966	6.8	81	22	055098
091up2 arabidopsis	413	6.8	80.5	23	091up2
009335 nosema locu	221	6.8	80	24	009335
026282 methanobact	402	6.8	80	25	026282
09vv18 drosophila	704	6.8	80	26	09vv18
030926 escherichia	821	6.8	80	27	030926
09rsb4 deinococcus	230	6.7	79	28	09RSB4
09sv13 arabidopsis	493	6.6	78.5	29	09SVJ3
09vsl1 drosophila	693	6.6	78.5	30	09VSL1
07690 bos taurus	1072	6.6	78.5	31	077690
029435 archaeoglob	248	6.6	78	32	029435
09yfi8 aeropyrum p	313	6.6	78	33	09YFI8
084200 chlamydia t	338	6.6	78	34	084200
066234 escherichia	704	6.6	78	35	066234
027560 methanobact	910	6.6	78	36	027560
Q9Z277 mus musculu	1479	6.6	78	37	Q9Z277
Q9xbn8 bacillus st	482	6.5	77.5	38	Q9XBN8
Q9Z020 mus musculu	754	6.5	77.5	39	Q9Z0R0
Q91j14 mus musculu	754	6.5	77.5	40	Q91J14
Q9p2r3 homo sapien	1166	6.5	77.5	41	Q9P2R3
Q9uya3 pyrococcus	253	6.5	77	42	Q9UYA3
Q9v977 drosophila	449	6.5	77	43	Q9V977
Q9z8d0 chlamydia p	659	6.5	77	44	Q9Z8D0
Q9jry4 chlamydia p	659	6.5	77	45	Q9JRY4

## ALIGNMENTS

RESULT 1  
Q9Y2Q3 PRELIMINARY; PRT; 226 AA.  
AC Q9Y2Q3;  
DT 01-NOV-1999 (TREMREL. 12, Created)  
DT 01-NOV-1999 (TREMREL. 12, Last sequence update)  
DT 01-NOV-1999 (TREMREL. 12, Last annotation update)  
DE GLUTATHIONE S-TRANSFERASE SUBUNIT 13 HOMOLOG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MAO M., Ye M., Zhang Q., Zhou J., Wu J., Shen Y., Kan L., He K.,  
RA Gu B., Fu G., Chen S., Chen Z.;  
RT "Human GSTK1-1 homolog gene."  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF070657; AAD20963.1;  
KW Transferase.  
SQ SEQUENCE 226 AA; 25497 MW; D3FDAFDI533B58A4 CRC64;

Query Match	100.0%;	Score 1185;	DB 4;	Length 226;
Best Local Similarity	100.0%;	Pred. No. 4.9e-96;		
Matches 226;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy 1	MGPLRTVELFDVLSFYSWGLGFEILCRYNQINWINLQRLPSLITGIMKDSGNKPPGLLP	60		
Db 1	MGPLRTVELFDVLSFYSWGLGFEILCRYNQINWINLQRLPSLITGIMKDSGNKPPGLLP	60		
Qy 61	RKGLYMANDLKLLRHHLQIPHFPKDFLSVMLEKGSLSAMRFLTAVNLEHPHMELEKASRE	120		
Db 61	RKGLYMANDLKLLRHHLQIPHFPKDFLSVMLEKGSLSAMRFLTAVNLEHPHMELEKASRE	120		
Qy 121	LMRYVSRNEDITEPOSILAAAEKAGNSAEOAQGLLEKIATPKVKNQLKETTAAACRYGA	180		
Db 121	LMRYVSRNEDITEPOSILAAAEKAGNSAEOAQGLLEKIATPKVKNQLKETTAAACRYGA	180		
Qy 181	FGLPITVAHVGDGTHMLFGSDRMELLAHLGKWMGPIPPAVNARL	226		
Db 181	FGLPITVAHVGDGTHMLFGSDRMELLAHLGKWMGPIPPAVNARL	226		

[illegible][illegible]

NCBI\_TaxID=6239;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favellio A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
RL Nature 0:0-0(0).  
[2]  
SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Favellio A.;  
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U13643; AAA21082.1; -;  
SQ SEQUENCE 312 AA; 35436 MW; 0FB8561F57687120 CRC64;

Query Match 12.3%; Score 145.5; DB 5; Length 312;  
Best Local Similarity 22.6%; Pred. No. 6.3e-05;  
Matches 51; Conservative 46; Mismatches 100; Indels 29; Gaps 6;  
QY 1 MGPLPRTVELFDVLSYPSW-----LGEFILCRYQ-----NIWNINLQ 38  
DB 1 MAPLPR-VKCYFDVCPNSWITTOITQALTSNLSFERIDFEPVCDKIGILHNAQIWNRRQ 59  
QY 39 LRPSLI--TGIMKDSGNKPPGLPRKGLYMAND---LKLRLHHLQIPHFPRKFLSVMLE 93  
DB 60 VHSNLSWTKTIEVPEQAESEETISENGILQKIDERGKLCICERVVPVDPKNTYKTAVA 119  
QY 94 KGSLSAMRFLTAVNLEHPEMLEKASRELWVRVSRNEDITEPQSILAAAEKAGMSAEQAO 153  
DB 120 RGSVPQLFLTSIREQYPOLYERAIHHLKRLQRLPQVHYGCHMSTVCRIGLISFKIAE 179  
QY 154 GLELIATPKVNQKLETTAEACRYGAFGLPITVAHD-GOTMLF 198  
DB 180 DIVARLSSPENRSILHKNCKEAVDFKLTEAPGLILLTDEGDTIKIF 225

Query Match 11.8%; Score 140; DB 2; Length 197;  
Best Local Similarity 23.1%; Pred. No. 0.0001;  
Matches 50; Conservative 40; Mismatches 94; Indels 32; Gaps 7;  
QY 4 LPRVELFDVLSYPSWGLFEILCRYQNIWNINLQRLPSLTITGIMKDSGNKPPG---LLP 50  
DB 1 MTRIDYFDISPSYLAQLKLPETARAAGCTVDYWPIDPEAKIRAGNYGSPNREVLP 60  
QY 61 RGLYMANDLKLRLHHLQIPHFPRKDF-----LSVMLEKGSLSAMRFLTAVNLEHPEML 114  
DB 61 -KIKVMKADLERWAERYGVPLTFPASAFACDMNCVLFAREHGKAEAVT-----109  
QY 115 EKASRELWVRVSRNEDITEPQSILAAAEKAGMSAEQAOGLLEKIATPKVNQKLETTAE 174  
DB 110 -----DAYRRIWGQIDPGRNELAAACATAAGL---DPAALIAFVESPAGONEYKRARSQ 161  
QY 175 ACRYGAFGLPITVAHDGQTHMLFGSDRMELLALL 210  
DB 162 AIQGVYCAPL--MEVDQ--IFWGNDRDLFLAEYL 193

RESULT 7  
Q92HH4 PRELIMINARY; PRT; 196 AA.  
ID Q92HH4  
AC Q92HH4  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)  
DE ISOMERASE PHND.  
GN PHND.  
OS Burkholderia sp. RP007.  
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
OC Burkholderia.  
OX NCBI\_TaxID=83784;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RP007.  
RX MEDLINE-99102215; PubMed-9882667;  
RA Laurie A.D., Lloyd-Jones G.;  
RT "The phn genes of Burkholderia sp. strain RP007 constitute a divergent  
RT gene cluster for polycyclic aromatic hydrocarbon catabolism."  
RL J. Bacteriol. 181:531-540(1999).  
DR EMBL; AF061751; AAD09871.1; -;  
KW isomerase.  
SQ SEQUENCE 196 AA; 21908 MW; 58716E9039BEA768 CRC64;

## RESULT 6

O85994  
ID O85994 PRELIMINARY; PRT; 197 AA.  
AC O85994;  
DT 01-NOV-1998 (Tremblrel. 08, Created)  
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
DE 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE.  
GN NAHD.  
OS Sphingomonas aromaticivorans.  
OC Plasmid pNL1.  
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;  
OC Sphingomonas.  
OX NCBI\_TaxID=48935;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-F199;  
RA Romine M.F., Stillwell L.C., Wong K.-K., Thurston S.J., Sisk E.C.,  
RA Sensen C.W., Gaasterland T., Saffer J.D., Fredrickson J.K.;  
RT "Complete sequence of a 184 kb catabolic plasmid from Sphingomonas  
RT aromaticivorans strain F199."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF079317; AAD04010.1; -;

Query Match 11.4%; Score 135; DB 2; Length 196;  
Best Local Similarity 24.0%; Pred. No. 0.00028;  
Matches 50; Conservative 35; Mismatches 103; Indels 20; Gaps 5;  
QY 7 TVELEYDLSYPSWGLFEILCRYQNIWNINLQRLPSLTITGIMKDSGNKPPGL--LPRKGL 64  
DB 2 TIDFFDFLSPYAYLARHRLTQVALHGCALAYKPIDLARAKLAIGNTGPANRDMPVKLA 61  
QY 65 YMANDLKLRLHHLQIPHFPRKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRELWVR 124  
DB 62 YVVEDLKRWAARYRIPIEFIKNTKRMVGT-----FYAEARGOQADYVROAYHIAWGE 116  
QY 125 VWSRNEDITEPQSILAAAEKAGMS-APQAGLLEKIATPKVNQKLETTAEACRYGAFGL 183  
DB 117 GGAPDD-----AALRSIAVSMGMDAADFLFLDSSEAEATYNSTLEAISAGVGV 168  
QY 184 PITVAHDGQTHMLFGSDRMELLALLIG 211  
DB 169 PTMAVGRD-----MWWGNDRIDFLLETHLG 192

RESULT 8



```
ID Q9WQX0 PRELIMINARY; PRT; 581 AA.
AC Q9WQX0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ABC TRANSPORTER, ATP-BINDING PROTEIN.
DE TM0043.
GN Thermotoga maritima.
OS Bacteria; Thermotogales; Thermotoga.
OC NCBI_TaxID=2336;
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Uitterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima."
DR EMBL: AE001691; AAD35137.1; --
DR HSSP: P13569; INBD.
DR TIGR: TM0043;
DR INTERPRO: IPR001140; --
DR INTERPRO: IPR001617; --
DR INTERPRO: IPR002106; --
DR PFAM: PF00005; ABC_tran; 1.
DR PFAM: PF00664; ABC_membrane; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR ATP-binding.
KW
SQ SEQUENCE 581 AA; 66324 MW; 55719E0DCA5D0A05 CRC64;

Query Match 7.7%; Score 91; DB 2; Length 581;
Best Local Similarity 19.0%; Pred. No. 8.4;
Matches 35; Conservative 44; Mismatches 67; Indels 38; Gaps 6;

QY 28 RYQNIW-----NINLQRLPSLITGIMKDSGNKPPGL-----LPRKGLYMAN 68
DB 343 RFENVWFSDYDGNVWLKDINLDFQPKLYAIVGTGGKSTLSLNGLYIPQKGNIFD 402
QY 69 DLKLRHHLQIP-----IHEPKDFL-----SVMLEKSGLSAMRFLTAVNLEHP-EMLE 115
DB 403 EIPLEYNLKLVRKQIAAYPQDVLFLSGTLDNIRLDESIPEERVLKARVHALDIE 462
QY 116 KASRELWVRWNRNEDITEPQSTLAAAEKAGM-----SAPQAQGLLEKIATPKVKNOLK 169
DB 463 RLPGVYVEIVERGTTLSAGERQLIARAVLFDKIFILDEATSNVDVITETKIQEALE 522
QY 170 ETTE 173
DB 523 ELSK 526

RESULT 12
ID O26322 PRELIMINARY; PRT; 410 AA.
AC O26322;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE FLAVOPROTEIN A HOMOLOG (II).
DE MT4220.
GN Methanobacterium thermoautotrophicum.
OS Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
OX NCBI_TaxID=2166;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=DELTA H.;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hwang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nollung J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RL deltaH: functional analysis and comparative genomics."
DR EMBL: AE000809; AAB84726.1; --
DR INTERPRO: IPR001279; --
DR PFAM: PF00753; lactamase_B; 1.
DR SEQUENCE 410 AA; 45740 MW; 8EAF9D3B363A8BD5 CRC64;

Query Match 7.6%; Score 90.5; DB 1; Length 410;
Best Local Similarity 25.1%; Pred. No. 5.8;
Matches 65; Conservative 31; Mismatches 82; Indels 81; Gaps 16;

QY 13 DVLSPSYWLGFELC-----RYQNIWNLQRLPSLITGIMKDSGNK----- 54
DB 99 EVLKRY---GSEIICTAKAEGRLROHYSIPQDTPMQ---TVKTGDDTLGKTLTFLEAP 152
QY 55 ----PPG---LLPRKGLYMANDLKLRHHLQIPHFPPKDFLSVMLEKSGLSAMRF-----L 103
DB 153 MLHWPDNSMFLLEEGILFSND--AFGQHLCLSKRFKDKVPEAVLMD---AAMKFYANLL 207
QY 104 TAVNLEHPMLEKAS--RELWM-----RVWSNRNEDITEPOSILAAAEKAGMSA 149
DB 208 TPLS---PLVLRKFSVEKELGELKIGMTAPSHGQIW-----TEPLKIIAA-----YT 252
QY 150 EQAOGILLEKIATPKVKNOLKETTAACRYGAFGLPITVAHVHVGDTMLFGSDRMELLHL 209
DB 253 DWATGKCRDKAT-IYDTMTHYSTRMLAHAMAEGL--MAADVDSMHFLHEDESEIVKNI 309
QY 210 LGKE-----WMGPIP 219
DB 310 LESKAVFTGSPFMFGPPF 328

RESULT 13
ID Q51499 PRELIMINARY; PRT; 199 AA.
AC Q51499;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE ISOMERASE.
DE PAHE.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAK1;
RA Takizawa N., Iida T., Yamauchi K., Satoh S., Wang Y., Fukuda M.,
RA Kiyohara H.;
RT "The molecular analysis of an NAH7-type gene cluster, pah, located on
RL the chromosome of Pseudomonas aeruginosa PAK1."
DR EMBL: D84146; BAA12247.1; --
KW Isomerase.
SQ SEQUENCE 199 AA; 22509 MW; 81C2A90CA56E5462 CRC64;

Query Match 7.5%; Score 89; DB 2; Length 199;
Best Local Similarity 24.2%; Pred. No. 3;
Matches 51; Conservative 37; Mismatches 95; Indels 28; Gaps 9;
```

QY 8 VLFYDVLSPYSLGFEILCYONINQININLQRLPSLITGIMKSGNKKPGL--LPKRGY 65  
 Db 3 VDFYDFLSPYSLGFEILCYONINQININLQRLPSLITGIMKSGNKKPGL--LPKRGY 62  
 QY 66 MANDLKLRRHLLQIPIHFPKDFLSVLEKSL--SAMRFLTA-VNLEHPEMLEKASRELW 122  
 Db 63 LKVDLQWAEYELIPIVFPANYNSRRNTGLYSGMAQTGAYNVV-----VF 110  
 QY 123 MRVSRNEDITEPOSILAA--AEKAGMSAEQAQGLLEKIAIPKVNOLKETTAAACRYGA 180  
 Db 111 NAVW--GGGIAPDLESPLVSEKLGWDRSAFE---DFISSDAATERDEQTHAALERKV 165  
 QY 181 FGLPTVAVHDGQTHMLFGSDRMELLALHLLG 211  
 Db 166 FGVP-TMFLGD---EMWNGNDRFLMLENAVG 192

RESULT 14  
 Q9SPF1  
 ID Q9SPF1 PRELIMINARY; PRT; 592 AA.  
 AC Q9SPF1;  
 DT 01-MAY-2000 (TREMELrel. 13, Created)  
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMELrel. 15, Last annotation update)  
 DE THREONINE DEHYDRATASE/DEAMINASE  
 GN OMRI.  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID-3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. GM11B;  
 RA Mourad G.S., Emerick R.M., Smith A.M.;  
 RT "Cloning and Sequencing of a cDNA Encoding an Isoleucine Feedback  
 RT Inhibitory Threonine Dehydratase/Deaminase of Arabidopsis thaliana  
 RT line GM11b (omri/omri).";  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF177212; AAD54324.1;  
 DR HSSP; P04968; 1TDJ.  
 DR INTERPRO; IPR000524;  
 DR INTERPRO; IPR000634;  
 DR INTERPRO; IPR001721;  
 DR INTERPRO; IPR001926;  
 DR PFAM; PF00291; PALP; 1.  
 DR PFAM; PF00585; Thr\_dehydrat\_C; 2.  
 DR PROSITE; PS00043; HTH\_GNTR\_FAMILY; UNKNOWN\_1.  
 DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
 SQ SEQUENCE 592 AA; 64562 MW; 1D02F73AD53FAP6C CRC64;

Query Match 7.5%; Score 89; DB 10; Length 592;  
 Best Local Similarity 22.8%; Pred. No. 13;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
 QY 55 PPGL-LPRKGLYMANDLKLRRHLLQIPIHFPKDFLSV-----LEKGSLS-AMRFLT-- 104  
 Db 52 PPKLPLPR-----LKVPNSLQ-----YPAGYLGAVPERTNEAENGSAEAMEYLNI 99  
 QY 105 -----AVNLEHP-EMLEKASRELWVRVSRNEDITE-----POSTLAA 141  
 Db 100 LSTKVYDIATESPLQAKLKLRRGLVRYLKREDLPQVPSFKLRGAYNMVVKLPADQJAK 159  
 QY 142 AEKAGMSAEQAQ-----LLEKIATPKVNOLKETTAAACRYGAFGLPTVA 188  
 Db 160 GVICSSAGNHAQGVALSASKLGCTAVIVMPVTTPEIKWQAVENL----- 203  
 QY 189 HVDGQTHMLFGSDRMELLALH---LGEKWMGP 220  
 Db 204 ---GATVVLFGDSYDQQAQAHAKIRAEERGLTFIPP 235

RESULT 15  
 Q9ZSS6  
 ID Q9ZSS6 PRELIMINARY; PRT; 592 AA.  
 AC Q9ZSS6;  
 DT 01-MAY-1999 (TREMELrel. 10, Created)  
 DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)  
 DT 01-OCT-2000 (TREMELrel. 15, Last annotation update)  
 DE THREONINE DEHYDRATASE/DEAMINASE (EC 4.2.1.16).  
 GN OMRI OR T22K18.12.  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID-3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Mourad G., Emerick R., Marion A., Smith A.;  
 RT "Cloning and Sequencing of a cDNA Encoding Threonine  
 RT Dehydratase/Deaminase of Arabidopsis thaliana (Accession No. AF096281)  
 RT (PCR 98-199).";  
 RL Plant Physiol. 118:1534-1534(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,  
 RA Ronning C.M., Koo H., Fujii C.Y., Otterback T.R., Barnstead M.E.,  
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
 RT "Arabidopsis thaliana chromosome III BAC T22K18 genomic sequence.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Mourad G.S., Smith A.M.;  
 RT "Molecular characterization of the genomic clone, including the  
 RT promoter sequences, of threonine dehydratase/deaminase from  
 RT Arabidopsis thaliana.";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF096281; AAC97936.1;  
 DR EMBL; AC010927; AAF04418.1;  
 DR EMBL; AF221984; AAF32370.1;  
 DR HSSP; P04968; 1TDJ.  
 DR INTERPRO; IPR000524;  
 DR INTERPRO; IPR000634;  
 DR INTERPRO; IPR001721;  
 DR INTERPRO; IPR001926;  
 DR PFAM; PF00291; PALP; 1.  
 DR PFAM; PF00585; Thr\_dehydrat\_C; 2.  
 DR PROSITE; PS00043; HTH\_GNTR\_FAMILY; UNKNOWN\_1.  
 DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
 KW Lyase.  
 SQ SEQUENCE 592 AA; 64634 MW; 16658747052FAE7C CRC64;

Query Match 7.5%; Score 89; DB 10; Length 592;  
 Best Local Similarity 22.8%; Pred. No. 13;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
 QY 55 PPGL-LPRKGLYMANDLKLRRHLLQIPIHFPKDFLSV-----LEKGSLS-AMRFLT-- 104  
 Db 52 PPKLPLPR-----LKVPNSLQ-----YPAGYLGAVPERTNEAENGSAEAMEYLNI 99  
 QY 105 -----AVNLEHP-EMLEKASRELWVRVSRNEDITE-----POSTLAA 141  
 Db 100 LSTKVYDIATESPLQAKLKLRRGLVRYLKREDLPQVPSFKLRGAYNMVVKLPADQJAK 159  
 QY 142 AEKAGMSAEQAQ-----LLEKIATPKVNOLKETTAAACRYGAFGLPTVA 188  
 Db 160 GVICSSAGNHAQGVALSASKLGCTAVIVMPVTTPEIKWQAVENL----- 203  
 QY 189 HVDGQTHMLFGSDRMELLALH---LGEKWMGP 220  
 Db 204 ---GATVVLFGDSYDQQAQAHAKIRAEERGLTFIPP 235

Search completed: February 15, 2001, 15:53:11  
Job time: 143 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2001, 15:50:42 ; Search time 16.3 Seconds  
(without alignments)  
474.098 Million cell updates/sec

Title: US-08-978-174-1

Perfect score: 1185

Sequence: 1 MGPLPRTVELFDVLSFYSW.....AHLGKWMGPIPPAVNARL.226

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_36.\*
- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.\*
  - 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.\*
  - 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.\*
  - 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.\*
  - 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.\*
  - 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.\*
  - 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.\*
  - 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.\*
  - 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.\*
  - 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.\*
  - 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.\*
  - 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.\*
  - 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.\*
  - 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.\*
  - 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.\*
  - 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.\*
  - 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.\*
  - 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.\*
  - 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.\*
  - 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.\*
  - 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1185	100.0	226	21	Novel human glutat
2	1185	100.0	256	20	Human endometrium
3	89	7.5	545	20	Mutant threonine d
4	89	7.5	545	20	Feedback insensiti
5	89	7.5	590	20	Mutant threonine d
6	89	7.5	592	20	Mutant threonine d
7	89	7.5	592	20	Mutant threonine d
8	89	7.5	592	20	Wild type threonin
9	89	7.5	592	20	Arabidopsis wild-t
10	89	7.5	592	20	Feedback insensiti
11	89	7.5	600	20	Mutant threonine d
12	89	7.5	609	20	Mutant threonine d

13	89	7.5	609	20	Y05704
14	83.5	7.0	308	17	W04266
15	83.5	7.0	539	20	Y32942
16	83.5	7.0	539	20	Y05706
17	81	6.8	911	20	Y55957
18	80.5	6.8	339	20	Y33825
19	80.5	6.8	532	20	Y32943
20	80.5	6.8	532	20	Y05707
21	79.5	6.7	1073	18	W32063
22	79.5	6.7	1073	19	W37371
23	78	6.6	341	20	W37006
24	77	6.5	659	20	Y35014
25	76.5	6.5	216	18	W24203
26	76.5	6.5	216	19	W55804
27	76.5	6.5	317	20	Y39338
28	76.5	6.5	317	20	Y23764
29	76.5	6.5	317	21	Y56814
30	76.5	6.5	497	13	R27786
31	76.5	6.5	497	16	R66893
32	76	6.4	588	20	Y30124
33	75.5	6.4	1398	17	R87008
34	75.5	6.4	1398	18	W24124
35	75.5	6.4	1398	20	W94839
36	74	6.2	342	21	Y91567
37	74	6.2	492	10	P94363
38	74	6.2	554	21	Y91569
39	74	6.2	912	20	Y55939
40	74	6.2	968	20	Y55966
41	73.5	6.2	869	15	R56501
42	73.5	6.2	869	17	W06091
43	73.5	6.2	869	18	W25022
44	73.5	6.2	980	20	W30811
45	73.5	6.2	985	20	W30607

## ALIGNMENTS

### RESULT 1

Y77499  
ID Y77499.standard; Protein; 226 AA.

AC Y77499;

DT 05-JUN-2000 (first entry)

DE Novel human glutathione S-transferase, GSTs.

KW Glutathione S-transferase; human; GSTs; cancer; immune disorder;  
KW gene therapy; diagnosis; treatment; drug screening.

OS Homo sapiens.

PN US6030809-A.

PD 29-FEB-2000.

PE 25-NOV-1997; 97US-0978174.

PR 25-NOV-1997; 97US-0978174.

PA (INCYTE), INCYTE PHARM INC.

PI Hillmap JL, Shah P, Lal P, Corley NC;

DR WPI; 2000-205204/18.

DR N-PSDB; 202599.

PT Isolated nucleic acid encoding glutathione S-transferase useful in the  
PT production of agents for preventing, diagnosing and treating diseases  
PT associated with cell proliferation -  
XX Claim 8; Fig 1A-C; 27pp; English.

Feedback insensiti  
Inosine-guanosine  
Mutant threonine d  
Feedback insensiti  
Mouse STE20-relate  
Amino acid sequenc  
Mutant threonine d  
Feedback insensiti  
Human ST receptor  
Human ST receptor  
Amino acid sequenc  
Chlamydia pneumoni  
Streptomyces non-m  
Streptomyces rosc  
Carboxymethyl cell  
A carboxymethyl ce  
T. maritima thermo  
Adrenodoxin reduct  
Human adrenodoxin  
A human protein wi  
Protease. Pyruvate  
Pyrococcus furiosu  
W09856926 Seq ID 6  
Human secreted pro  
Rovine adrenal gla  
Human secreted pro  
Human GSK2 protein  
Full length human  
TATA-binding prote  
Human TATA-binding  
TATA-binding prote  
Arabidopsis thalia  
Arabidopsis thalia

XX This sequence represents a novel human glutathione S-transferase, GSTS.  
CC Nucleotide sequences encoding GSTS were initially isolated from a  
CC urologic cDNA library, and subsequently extended using cDNA libraries  
CC derived from other tissues, such as brain or bladder. The present  
CC sequence is encoded by a consensus cDNA, GSTS, and nucleotides which  
CC encode it may be used in the prevention, treatment and diagnosis of  
CC diseases associated with inappropriate GSTS expression, such as cancers  
CC and immune disorders. Nucleotides which encode GSTS may be used in gene  
CC therapy to treat disorders associated with reduced expression or activity  
CC of GSTS, and in antisense therapy for disorders associated with increased  
CC GSTS expression or activity. They may also be used for the recombinant  
CC production of GSTS, and as a source of probes and primers to detect and  
CC quantitate the presence of similar nucleic acid sequences, particularly  
CC for the diagnosis of GSTS-associated disorders. GSTS proteins may be used  
CC as antigens in the production of antibodies against GSTS and in assays to  
CC identify modulators (agonists and antagonists) of GSTS expression and  
CC activity. The anti-GSTS antibodies and GSTS antagonists may also be used  
CC to downregulate GSTS expression and activity. Antagonists of GSTS  
CC expression and function may be used to treat immune disorders (e.g., AIDS,  
CC anaemia, asthma, Crohn's disease, irritable bowel syndrome, multiple  
CC sclerosis, osteoarthritis, microbial infections) and cancers (e.g.,  
CC leukaemia, lymphoma, melanoma, and cancers of the breast, prostate,  
CC liver, lung and brain). The anti-GSTS antibodies may also be used as  
CC diagnostic agents.

XX Sequence 226 AA;  
SQ Query Match 100.0%; Score 1185; DB 21; Length 226;  
Best Local Similarity 100.0%; Pred. No. 4.5e-122;  
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPIPTVELFYDVLSPYSWLGFEILCRYQNTWINQLRPSLTGIMKDSGNKPPGLLP 60  
Db 1 mgp1prtvelfydvlspyswlgfeilcryqntwinqlrpsltgimkdsngkppgllp 60  
QY 61 RKGLYMANDLKLRLHQLPIHFPHKDFLSVMLEKSGLSAMRFLTAVNLEHPEMLEKASRE 120  
Db 61 rkglymandlklrlhqlpilhfpkdfslvmleksglsamrfltavnllehpeMLEKASRE 120  
QY 121 LWMRVWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQLKETTEAACRYGA 180  
Db 121 lwmrvwsrneditepqsilaaekagmsaeqaglllekiaatkpkvknqlketteaacyga 180  
QY 181 FGLPITVAHVDCQTHMLFGSDRMELLALLHLLGEKWMGPPIPPAVNARL 226  
Db 181 fg1pitvahvdgqthmlfgsdrnellahllhllgekwmgppipavnar1 226

RESULT 2  
Y59988  
ID Y59988 standard; Protein: 256 AA.  
XX Y59988;  
AC AC  
XX 31-JAN-2000 (first entry)  
DT DT  
XX Human endometrium tumour EST encoded protein 48.  
DE DE  
XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:  
KW treatment; uterine; gene therapy; expressed sequence tag.  
XX Homo sapiens.  
OS DE19817948-A1.  
XX 21-OCT-1999.  
PD 17-APR-1998; 98DE-1017948.  
PF 17-APR-1998; 98DE-1017948.  
XX 17-APR-1998; 98DE-1017948.  
PR 17-APR-1998; 98DE-1017948.  
XX

(META-) METAGEN-GES GENOMFORSCHUNG MBH.

PA Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;  
XX WPT; 1999-591957/51.  
PI N-PSDB: 241996.  
DR New nucleic acid sequences expressed in uterine cancer tissues, and  
XX derived polypeptides, for treatment of uterine and endometrial cancer  
XX and identification of therapeutic agents -  
PS Claim 23; Page 294; 444pp; German.

XX This invention describes novel human nucleic acid (cDNA) sequences (A),  
CC that are highly expressed in uterine tumour tissue and which have  
CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
CC are used (i) to identify agents suitable for treatment of uterine or  
CC endometrial cancer; (ii) directly for treating these forms of cancer  
CC (including expression from gene therapy vectors) and (iii) for generation  
CC of specific antibodies. (A) are identified by assembling ESTs (expressed  
CC sequence tags) from a particular tissue type before comparison of the  
CC expression patterns. This allows a significantly longer fragment of the  
CC gene to be revealed, so should reduce the number of failures associated  
CC with the fact that ESTs from different libraries may represent different  
CC parts of the same unknown gene, distorting the estimated frequency of  
CC occurrence in a particular tissue. Y59941-Y60328 represent protein  
CC fragments encoded by the human endometrium tumour cDNA library derived  
CC EST fragments represented in 241981-242121.

XX Sequence 256 AA;

Query Match 100.0%; Score 1185; DB 20; Length 256;  
Best Local Similarity 100.0%; Pred. No. 5.4e-122;  
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPIPTVELFYDVLSPYSWLGFEILCRYQNTWINQLRPSLTGIMKDSGNKPPGLLP 60  
Db 31 mgp1prtvelfydvlspyswlgfeilcryqntwinqlrpsltgimkdsngkppgllp 90  
QY 61 RKGLYMANDLKLRLHQLPIHFPHKDFLSVMLEKSGLSAMRFLTAVNLEHPEMLEKASRE 120  
Db 91 rkglymandlklrlhqlpilhfpkdfslvmleksglsamrfltavnllehpeMLEKASRE 150  
QY 121 LWMRVWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQLKETTEAACRYGA 180  
Db 151 lwmrvwsrneditepqsilaaekagmsaeqaglllekiaatkpkvknqlketteaacyga 210  
QY 181 FGLPITVAHVDCQTHMLFGSDRMELLALLHLLGEKWMGPPIPPAVNARL 226  
Db 211 fg1pitvahvdgqthmlfgsdrnellahllhllgekwmgppipavnar1 256

RESULT 3  
Y32947  
ID Y32947 standard; Protein: 545 AA.  
XX Y32947;  
AC AC  
XX 09-NOV-1999 (first entry)  
DT DT  
XX Mutant threonine dehydratase/deaminase protein sequence.  
DE DE  
XX Threonine dehydratase/deaminase; TD: feedback insensitive mutant;  
KW molecular marker; isoleucine toxic structural analog resistance;  
KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
KW polyhydroxybutyrate; antibiotic resistance marker; muten.  
XX Arabidopsis thaliana.  
OS Synthetic.  
XX WO9941395-A1.  
PN



QY 189 HVDGTHMLFGSDRMELLAHL---LGEKWMGPIPP 220  
 ID Y32950 standard; Protein; 590 AA.  
 AC Y32950;  
 XX  
 XX  
 DT 09-NOV-1999 (first entry)  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 DE Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutin.  
 XX Arabidopsis thaliana.  
 OS Synthetic.  
 OS W09941395-A1.  
 PN 19-AUG-1999.  
 PD 08-JAN-1999; 99WO-US00560.  
 PE 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; Z11208.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Example 3; Page 119-123; 194pp; English.

QY 189 HVDGTHMLFGSDRMELLAHL---LGEKWMGPIPP 220  
 ID Y32950 standard; Protein; 590 AA.  
 AC Y32950;  
 XX  
 XX  
 DT 09-NOV-1999 (first entry)  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 DE Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutin.  
 XX Arabidopsis thaliana.  
 OS Synthetic.  
 OS W09941395-A1.  
 PN 19-AUG-1999.  
 PD 08-JAN-1999; 99WO-US00560.  
 PE 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; Z11208.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Example 3; Page 119-123; 194pp; English.

QY 55 PPGL-LPRKGLYMANDLLRHLHLPKFLSVW-----LEKGLS-AMRFLT-- 104  
 DB 50 ppklplpr-----lkvpsnslq---ypagylgavpntneangsaameyltni 97  
 QY 105 -----AVNLEHP-EMLEKASRELWVRVNSRNEDITE-----PQSILAA 141

Db 98 lskvdydiaesplqlaklksrlgvmlykredlqpvfsfklrgaynmvklpadqlak 157  
 QY 142 AERAGMSAEQHQG-----LLEKIATPKVKNQKLTETEAACRYGAFGLPITVA 188  
 Db 158 gvicssaghuqgvalsaskigctavivmpvtteikwqavenl----- 201  
 QY 189 HVDGTHMLFGSDRMELLAHL---LGEKWMGPIPP 220  
 Db 202 ---gatvvlfgdsydaqahakiraeegitfipp 233

RESULT 6  
 Y32939  
 ID Y32939 standard; Protein; 592 AA.  
 XX Y32939;  
 AC  
 XX  
 DT 09-NOV-1999 (first entry)  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 DE Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutin.  
 XX Arabidopsis thaliana.  
 OS Synthetic.  
 OS W09941395-A1.  
 PN 19-AUG-1999.  
 PD 08-JAN-1999; 99WO-US00560.  
 PE 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; Z11197.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Claim 13; Fig 8; 194pp; English.

QY 55 PPGL-LPRKGLYMANDLLRHLHLPKFLSVW-----LEKGLS-AMRFLT-- 104  
 DB 50 ppklplpr-----lkvpsnslq---ypagylgavpntneangsaameyltni 97  
 QY 105 -----AVNLEHP-EMLEKASRELWVRVNSRNEDITE-----PQSILAA 141

Query Match 7.5%; Score 89; DB 20; Length 592;  
 Best Local Similarity 22.8%; Pred. No. 0.34; Mismatches 29; Indels 80; Gaps 11;  
 Matches 49; Conservative 29; Mismatches 29; Indels 80; Gaps 11;

us-08-978-174-1.rag

Thu Feb 15 15:59:12 2001

SQ Sequence 592 AA;  
 Query Match 7.5%; Score 89; DB 20; Length 592;  
 Best Local Similarity 22.8%; Pred. No. 0.34;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
 QY 55 PPGL-LPRKGLYMANDLKLRLHLLQIPHFPPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
 Db 52 ppklplpr-----lkvspnslq-----ypagylgavperrneangsiacameyltnl 99  
 QY 105 -----AVNLEHP-EMLEKASRELWVRVSRNEDITE-----PQSILAA 141  
 Db 100 lskkydiaiesplqlakklslrgvmylkredlqpvfslrgaynmvklpadqlak 159  
 QY 142 AEKAGMSAEQAQG-----LLEKIATPKVNQKLTETTEACRYGAFGLPIIVA 188  
 Db 160 gvicssagnhagvalsasklgtavimpvttpeikwqvenl----- 203  
 QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220  
 Db 204 ---gatvvlfgdsydaqahakiraeegltfipp 235  
 RESULT 8  
 Y32951  
 ID Y32951 standard; Protein; 592 AA.  
 XX AC Y32951;  
 XX 09-NOV-1999 (first entry)  
 DT Wild type threonine dehydratase/deaminase protein sequence.  
 DE Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 XX molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker.  
 XX Arabidopsis thaliana.  
 OS Arabidopsis thaliana.  
 XX WO9941395-A1.  
 XX 19-AUG-1999.  
 XX 08-JAN-1999; 99WO-US00560.  
 XX 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX (DOWC) DOW AGROSCIENCES LLC.  
 PA (PURD) PURDUE RES FOUND.  
 XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 WPI; 1999-527375/44.  
 DR N-PSDB; Z11209.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Disclosure; Page 86-89; 194pp; English.  
 XX This sequence is the wild type Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein. The invention relates to mutants of  
 CC the encoded protein, that are feedback insensitive TD mutants. The TD DNA  
 CC sequence is used as molecular marker (impacting resistance to toxic  
 CC structural analogues of isoleucine) for selecting transformed cells and  
 CC to produce transformants with increased levels of isoleucine (and thus  
 CC better nutritional value) or of intermediates in biosynthesis of  
 CC isoleucine (e.g. 2-oxobutyrate, for synthesis of the degradable  
 CC biopolymer poly(hydroxybutyrate)). Also TD-expressing plants permit use

QY 55 PPGL-LPRKGLYMANDLKLRLHLLQIPHFPPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
 Db 52 ppklplpr-----lkvspnslq-----ypagylgavperrneangsiacameyltnl 99  
 QY 105 -----AVNLEHP-EMLEKASRELWVRVSRNEDITE-----PQSILAA 141  
 Db 100 lskkydiaiesplqlakklslrgvmylkredlqpvfslrgaynmvklpadqlak 159  
 QY 142 AEKAGMSAEQAQG-----LLEKIATPKVNQKLTETTEACRYGAFGLPIIVA 188  
 Db 160 gvicssagnhagvalsasklgtavimpvttpeikwqvenl----- 203  
 QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220  
 Db 204 ---gatvvlfgdsydaqahakiraeegltfipp 235  
 RESULT 7  
 Y32948  
 ID Y32948 standard; Protein; 592 AA.  
 XX AC Y32948;  
 XX 09-NOV-1999 (first entry)  
 DT Mutant threonine dehydratase/deaminase protein sequence.  
 DE Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 XX molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutcin.  
 XX Arabidopsis thaliana.  
 OS Synthetic.  
 XX WO9941395-A1.  
 XX 19-AUG-1999.  
 XX 08-JAN-1999; 99WO-US00560.  
 XX 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX (DOWC) DOW AGROSCIENCES LLC.  
 PA (PURD) PURDUE RES FOUND.  
 XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 WPI; 1999-527375/44.  
 DR N-PSDB; Z11206.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Example 3; Page 126-129; 194pp; English.  
 XX This sequence represents a mutant Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 CC marker (impacting resistance to toxic structural analogues of isoleucine)  
 CC for selecting transformed cells and to produce transformants with  
 CC increased levels of isoleucine (and thus better nutritional value) or of  
 CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate; for  
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
 CC TD-expressing plants permit use of the isoleucine structural analogues as  
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
 CC markers (which are potentially harmful to the environment). Since no  
 CC human analog of TD exists (humans can not synthesize isoleucine), it  
 CC should be safe to use.



```

Region      536..554
            /note= "regulatory region R6"
Misc-difference 499
            /note= "Arg in wild-type TD"
Misc-difference 544
            /note= "Arg in wild-type TD"
WO9902656-AI.
21-JAN-1999.
10-JUL-1998; 98WO-US14362.
17-FEB-1998; 98US-0074875.
10-JUL-1997; 97US-0052096.
(PURD ) PURDUE RES FOUND.
Mourad GS;
WPI; 1999-120860/10.
N-PSDB; X25332.
New sequences encode mutant threonine dehydratase/deaminase - which
is insensitive to feedback inhibition, useful as a selective marker
to produce transformed cells resistant to toxic isoleucine analogues
disclosure; Page 53-56; 120pp; English.
The present sequence represents an Arabidopsis thaliana mutant
threonine dehydratase/deaminase (TD) precursor which, unlike
wild-type TD, is insensitive to feedback inhibition by isoleucine.
Claimed polynucleotides (see X25332-40), originally isolated and
cloned from A. thaliana mutated line GM1b (omrl/omrl), encode
feedback insensitive TD that can be used to transform a wide
variety of plants, fungi, bacteria and yeast. Mutant TD differs
from the wild-type enzyme only by an R499C amino acid substitution
in regulatory region R4, and by an R544H substitution in regulatory
region R6. Mutant TD is not only insensitive to feedback
inhibition by isoleucine, but is also insensitive to structural
analogues of isoleucine that are toxic to plants and microorganisms
which synthesize only wild-type TD. Nucleotide sequences encoding
mutated forms of TD can therefore be used to create cells that are
insensitive to compounds normally toxic to cells expressing only
wild-type TD enzymes, and thus may be used to provide a biochemical
selectable marker. Transformsants that express the mutant TD show
increased levels of isoleucine production, and thus provide an
improved nutrient source.
Sequence 592 AA;
Query Match      7.5%; Score 89; DB 20; Length 592;
Best Local Similarity 22.8%; Pred. No. 0.34;
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;
QY 55 BPGL-LPRGLXYMANDLKLRLHRLQIPHFPKDFLSVM-----LKGSL-S-AMRFLT-- 104
DB 52 ppklplpr-----lkvpsnlq---ypagylgavpertaineangslaemeyltni' 99
QY 105 -----AVNLEHP-EMLEKASRELMRMVRSNRNDETE-----PQSLIAA 141
DB 100 lsktvvdiatesplqlakklslkgrvmylkredlqvpsfklrgaynmvklpadqlak 159
QY 142 AEKAGNSAEQAQG-----LLEKIATPKVKNLKETTEAACRYGAFGLPTTVA 188
DB 160 gvicsaagnhagdvalsasklgctavimpvtbtpelkwgavnl----- 203
QY 189 HVDGQTHLFGSDRMELLAHL'---LGEKWMGPIPP 220
DB 204 ---gatvvlfgdsydaqahakiraeegltfipp 235

```





CC structural analogues of isoleucine that are toxic to plants and  
 CC microorganisms which synthesize only wild-type TD. Nucleotide  
 CC sequences encoding mutated forms of TD can therefore be used to  
 CC create cells that are insensitive to compounds normally toxic to  
 CC cells expressing only wild-type TD enzymes, and thus may be used  
 CC to provide a biochemical selectable marker. Transformants  
 CC harboring a nucleotide sequence comprising a promoter operably  
 CC linked to a mutated TD sequence demonstrate increased levels of  
 CC isoleucine production, and thus provide an improved nutrient source.  
 XX  
 SQ Sequence 609 AA;

Query Match 7.5%; Score 89; DB 20; Length 609;  
 Best Local Similarity 22.8%; Pred. NO. 0.36; 57; Indels 80; Gaps 11;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

QY 55 PPGL-LPRKGLYANDKLLRHHLQIPHPKDFLSVM-----LEKGLS-AMRFLT-- 104  
 DB 69 pklplpr-----lkvspnslq---ypagylgavperrneangsiameylni 116  
 QY 105 -----AVNLEHP-EMLEKASRELMMRWVSRNEDITE-----POSILAA 141  
 DB 117 lskvydiaiesplqlaklslrgvmyikredlqpvsfklrgaynmwklpadqlak 176  
 QY 142 AEKAGMSAEQAQ-----LLEKIATPKVNQLKETTAAACRYGAFGLPIITVA 188  
 DB 177 gvicssagnhagdvaisaklgtctavimpvttpeikwqvenl----- 220  
 QY 189 HVDGOTMLFGSDRMELLALH---LGEKWMGPIPP 220  
 DB 221 ---gatvlfsgsydaqahakiraeegitfipp 252

RESULT 14  
 W04266  
 ID W04266 standard; Protein; 308 AA.  
 AC W04266;  
 XX 30-JUN-1997 (first entry)  
 XX Inosine-guanosine kinase.  
 XX Inosine-guanosine kinase; 5'-inosinic acid; 5'-guanylic acid;  
 KW Corynebacterium ammoniagenes; ATP; seasoning; food.  
 XX Exiguobacterium sp. (ATCC 35652).  
 XX W09630501-A1.  
 XX 03-OCT-1996.  
 XX 22-MAR-1996; 96WO-JP00761.  
 XX 09-JUN-1995; 95JP-0177900.  
 XX 24-MAR-1995; 95JP-0102888.  
 XX (AJIN ) AJINOMOTO CO INC.  
 XX Kawasaki H, Shimaoka M, Usuda Y, Utagawa T;  
 PI WPI: 1996-455349/45.  
 DR N-PSDB; T33972.  
 XX Prodn. of 5'-inosinic acid or 5'-guanylic acid - using C.  
 PT ammoniagenes transformed by inosine-guanosine kinase gene, for use,  
 PT in food seasoning  
 XX Example 14; Page 65-66; 72pp; Japanese.  
 PS This sequence represents the inosine-guanosine kinase sequence derived  
 CC from Exiguobacterium sp. This sequence was used in the production of

CC 5'-inosinic acid or 5'-guanylic acid by reacting an organism, pref.  
 CC Corynebacterium ammoniagenes, capable of regenerating ATP, containing  
 CC this recombinant DNA, with inosine, guanosine or their precursors, and  
 CC an energy and phosphoric acid source. 5'-inosinic acid and 5'-guanylic  
 CC acid are useful as seasonings for food.  
 XX  
 SQ Sequence 308 AA;

Query Match 7.0%; Score 83.5; DB 17; Length 308;  
 Best Local Similarity 25.8%; Pred. NO. 0.52; 57; Indels 33; Gaps 8;  
 Matches 40; Conservative 25; Mismatches 57; Indels 33; Gaps 8;

QY 63 GLYWANDLKLRLHLLQIPHPKDFLSVMLEKGLS-AMRFLTAVNLEHPEMLEKASRELW 122  
 DB 40 grnvaqlgvlgnvdrfvstvtndqig-----gvleelrsln-vnvehvdilledngmgw 94  
 QY 123 MRVMSRNEDITEPPQSILAAAEKAGMSAEQAQGLLEKATPKVK-----NOLKE 170  
 DB 95 lavmdnngdl--qtsiskqpdeamme---qcilrridvtfaestavaldidlsnvlne 148  
 QY 171 TTEACR-----YGAFLGLPIITVAHVVDGQTHMLFG 199  
 DB 149 tie-lcremkpilygvvcg---hlsviernrhliqg 179

RESULT 15  
 Y32942  
 ID Y32942 standard; Protein; 539 AA.  
 XX Y32942;  
 XX 09-NOV-1999 (first entry)  
 XX Mutant threonine dehydratase/deaminase protein sequence.

Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; muten.  
 XX Arabidopsis thaliana.  
 OS Synthetic.

XX W09941395-A1.  
 XX 19-AUG-1999.  
 XX 08-JAN-1999; 99WO-US00560.  
 XX 10-JUL-1998; 98WO-US14362.  
 XX 17-FEB-1998; 98US-0074875.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 XX (PURD ) PURDUE RES FOUND.  
 XX Larrinua IM, Merlo DJ, Mourad GS, Paredy DR;  
 DR WPI: 1999-527375/44.  
 DR N-PSDB; Z11200.

New nucleic acid encoding threonine dehydratase deaminase resistant  
 to feedback inhibition, useful as selection marker for cell  
 transformation and to impart herbicide resistance  
 PS Claim 13; Page 106-109; 194pp; English.

This sequence represents a mutant Arabidopsis thaliana threonine  
 dehydratase/deaminase (TD) protein of the invention. The protein is a  
 feedback insensitive mutant. The TD DNA sequence is used as molecular  
 marker (imparting resistance to toxic structural analogues of isoleucine)  
 for selecting transformed cells and to produce transformants with  
 increased levels of isoleucine (and thus better nutritional value) or of

CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for  
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
 CC TD-expressing plants permit use of the isoleucine structural analogues as  
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
 CC markers (which are potentially harmful to the environment). Since no  
 CC human analog of TD exists (humans can not synthesize isoleucine), it  
 CC should be safe to use.  
 XX  
 SQ Sequence 539 AA;

Query Match 7.0%; Score 83.5; DB 20; Length 539;  
 Best Local Similarity 21.9%; Pred. No. 1.2;  
 Matches 46; Conservative 29; Mismatches 56; Indels 79; Gaps 10;  
 QY 59 LPRKGLYMANDLLRHHHQIPIHFKDPLSVN-----LEKGSLS-AMRFLT----- 104  
 Db 4 lpr-----lkvspuslg-----ypagylgavperrneaeagsiaeameyitnllstkv 51  
 QY 105 -AVNLEHP-EMLEKASRELWVRVMSRNEDETE-----PQSILAAAEKAG 146  
 Db 52 ydiaiesplqlaklksrlgvrmylkredlcpvfsklrgaynmvmvklpadqlakgvics 111  
 QY 147 MSAEQHQG-----LLEKIATPKVKNQKLTETEAACRYGAFGLPITVAHVDDQ 193  
 Db 112 sagnhaqgvalsasklgctavimpvttpeikqavenl-----ga 152  
 QY 194 THMLFGSDRMELLAHL---LGEKWMGPIPP 220  
 Db 153 tvvlfqdsydaqahakiraeegitfipp 182

Search completed: February 15, 2001, 15:52:05  
 Job time: 83 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2001, 15:50:48 ; Search time 14.15 Seconds  
(without alignments)  
286.805 Million cell updates/sec

Title: US-08-978-174-1  
Perfect score: 1195  
Sequence: 1 MGPLPRTVELFDVLSYV.....AHLGKWMGPPIPAVNRL 226

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues  
Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1185	100.0	226	3	US-08-978-174-1
2	874	73.8	226	3	US-08-978-174-3
3	80.5	6.8	339	2	US-08-855-714-3
4	76.5	6.5	317	2	US-09-066-075-2
5	76.5	6.5	317	2	US-08-518-615A-2
6	76.5	6.5	317	3	US-08-951-889-2
7	76.5	6.5	497	1	US-08-075-193-4
8	76.5	6.5	497	2	US-08-564-090A-4
9	76.5	6.5	497	4	PCT-US94-06698-4
10	75.5	6.4	1398	1	US-08-750-532-9
11	73.5	6.2	523	2	US-08-473-553A-3
12	73.5	6.2	869	1	US-08-188-582-32
13	73.5	6.2	869	1	US-08-646-715-32
14	73.5	6.2	980	2	US-08-473-553A-6
15	73.5	6.2	985	2	US-08-473-553A-2
16	73	6.2	948	1	US-08-698-551-14
17	73	6.2	948	2	US-08-602-228-14
18	73	6.2	948	2	US-08-533-901B-14
19	73	6.2	948	2	US-08-839-032A-14
20	73	6.2	948	2	US-08-839-031A-14
21	73	6.2	948	4	PCT-US95-12724-14
22	71.5	6.0	543	2	US-08-922-170B-10
23	71	6.0	587	1	US-07-955-905A-23
24	70.5	5.9	615	2	US-08-484-101B-38
25	70	5.9	401	1	US-08-198-446B-11
26	70	5.9	401	2	US-08-870-693-11
27	69.5	5.9	259	2	US-07-857-224B-51
28	69.5	5.9	529	4	PCT-US95-0500B-15

Sequence 2, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 6, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 10, Appli  
Sequence 50, Appli  
Sequence 14, Appli  
Sequence 4, Appli  
Sequence 4, Appli

## ALIGNMENTS

RESULT 1  
US-08-978-174-1  
; Sequence 1, Application US/08978174  
; Patent No. 6030809  
; GENERAL INFORMATION:  
; APPLICANT: Shah, Purvi  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: NEW GLUTATHIONE-S-TRANSFERASE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,174  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0430 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 226 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BLADTUT04  
; CLONE: 1554593  
US-08-978-174-1

Query Match

100.0% Score 1185; DB 3; Length 226;

```

Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

\_\_\_\_\_

10

1

Db 47 PLPTSPMLKMCQSAPPTSWLKRRLFCPPRPHWK-----SLRTASQKSSFSSTKAALK 99  
Qy 60 PRK-GLYMANDLKLRLHQLQIPHFDPKDFSLVMLEKGS-----SAMRELTAVNLH 110  
Db 100 PRPGSWTRKSSPSRSLGRANH-FODRLGIDGKGVRLASLDETQACNAFAAIN-KA 157  
Qy 111 PEMLE---KASRELWMRVMSRNETITEPQSILAAAKAG-----TPKVNQLKETTE 173  
Db 158 PAILEGFVEFEV-----SVIAARDRSGNVAIFDLAENVHKGILATST 202  
Qy 147 -----MSAEQAQGLLEKTA-----TPKVNQLKETTE 173  
Db 203 VPAASVQTAEAARTAAEKHLHALDYGVLGFEFVLKDGTLTLANEFAPRVHNS-GHWTE 261  
Qy 174 AACRYGAF-----GLPT--TVAHVDGOTHMLFGSD-----RMELLAHLGK 213  
Db 262 AACATSOFEQHTRAVAGLPLGNTDRHSDCVNENLIGDIEKVPAILCEKNVHLHYGKK 320

RESULT 4  
US-09-066-075-2  
Sequence 2, Application US/09066075  
Patent No. 5925749  
GENERAL INFORMATION:  
APPLICANT: Mathur, E., et al.  
TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,075  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/518,615  
FILING DATE: August 23, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 331400-20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 317 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-09-066-075-2

Query Match 6.5%; Score 76.5; DB 2; Length 317;  
Best Local Similarity 21.1%; Pred. No. 1.5;  
Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;  
Qy 76 HLQIPIH-----FP-----KDFLSVMLEKGSLSAMRFLTAVNLHPHPEM 116  
Db 49 HVRIPRWSTHAYAFPPYKIMDRFFKRVDEVINGALKRG-----LAVAINIHHELMN 102  
Qy 117 ASRE-----LWMRVMSRNE-----ITEPQS-----137

Db 103 DPEEHKERFLALWKQIADRYKDYPTLFFELLNPHGNLTPEKWNELLEALKVIRSIDK 162  
Qy 138 ----TLAAAEKAGMSAEQAQGLLEKTIATPKVNQLKETTEACRYGATGLPITVAHVDGQ 193  
Db 163 KHTIIIGTAEWGGISA-----LEKUSVPKWE---KNSIVTIHYNPFEF-----203  
Qy 194 THMLFGSDRMELLAHLGKWMGP 217  
Db 204 THQ--GAWEVEGSEKWLGRKWGSP 225

RESULT 5  
US-08-518-615A-2  
Sequence 2, Application US/08518615A  
Patent No. 5962258  
GENERAL INFORMATION:  
APPLICANT: Mathur, E., et al.  
TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/518,615A  
FILING DATE: August 23, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 331400-20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 317 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-518-615A-2

Query Match 6.5%; Score 76.5; DB 2; Length 317;  
Best Local Similarity 21.1%; Pred. No. 1.5;  
Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;  
Qy 76 HLQIPIH-----FP-----KDFLSVMLEKGSLSAMRFLTAVNLHPHPEM 116  
Db 49 HVRIPRWSTHAYAFPPYKIMDRFFKRVDEVINGALKRG-----LAVAINIHHELMN 102  
Qy 117 ASRE-----LWMRVMSRNE-----ITEPQS-----137  
Db 103 DPEEHKERFLALWKQIADRYKDYPTLFFELLNPHGNLTPEKWNELLEALKVIRSIDK 162  
Qy 138 ----TLAAAEKAGMSAEQAQGLLEKTIATPKVNQLKETTEACRYGATGLPITVAHVDGQ 193  
Db 163 KHTIIIGTAEWGGISA-----LEKLSVPKWE---KNSIVTIHYNPFEF-----203



STREET: 3000 EL CAMINO REAL  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/364,090A  
FILING DATE: 02/05/96  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD L. NEELEY, PH.D.  
REGISTRATION NUMBER: 30,092  
REFERENCE/DOCKET NUMBER: UCAL-236/01US  
TELEPHONE: 415-843-5000  
TELEFAX: 415-857-0663  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 497 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-564-090A-4

Query Match 6.5%; Score 76.5; DB 2; Length 497;  
Best Local Similarity 25.4%; Pred. No. 3;  
Matches 44; Conservative 29; Mismatches 69; Indels 31; Gaps 9;  
QY 14 VLSPYSVLGFEILCRYQNIWNINLQRLPSLITGIMKDSGNKPPGLPRKG-LYMANDLKL 72  
DB 195 LLTPPEHLEALLCQRTDITKAAL-----GVLRQSRVKTVMVVGRRGPLQVAFITKE 246  
QY 73 LRHLQI-----PIHPKDFL-----SYMLEKGSLSAMRFLAVNLEHEPMELEKASREL 121  
DB 247 LREMIQPGARPILDVDFGLQDKIKEVPRPKRTELLLRAT--EKPGPAEAARQAS 304  
QY 122 WMRVSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKYKNQKTEEA 174  
DB 305 ASRAWGL-REFRSPQVLPSPD-----GRRAGV--RLAVTRLEG-VDEATRA 348

RESULT 9  
PCT-US94-06698-4  
Sequence 4, Application PC/TUS9406698  
GENERAL INFORMATION:  
APPLICANT: MILLER, WALTER L.  
APPLICANT: HARIKRISHNA, JENNIFER A.  
TITLE OF INVENTION: CHOLESTEROL DISPOSAL FUSION ENZYMES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ROBBINS, BERLINER & CARSON  
STREET: 201 NORTH FIGUEROA STREET  
CITY: LOS ANGELES  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 90012  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/06698  
FILING DATE: FILED HERewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: BERLINER, ROBERT  
REGISTRATION NUMBER: 20,121  
REFERENCE/DOCKET NUMBER: 5555-224-C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 213-977-1001  
TELEFAX: 213-977-1003  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 497 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-06698-4

Query Match 6.5%; Score 76.5; DB 4; Length 497;  
Best Local Similarity 25.4%; Pred. No. 3;  
Matches 44; Conservative 29; Mismatches 69; Indels 31; Gaps 9;  
QY 14 VLSPYSVLGFEILCRYQNIWNINLQRLPSLITGIMKDSGNKPPGLPRKG-LYMANDLKL 72  
DB 195 LLTPPEHLEALLCQRTDITKAAL-----GVLRQSRVKTVMVVGRRGPLQVAFITKE 246  
QY 73 LRHLQI-----PIHPKDFL-----SYMLEKGSLSAMRFLAVNLEHEPMELEKASREL 121  
DB 247 LREMIQPGARPILDVDFGLQDKIKEVPRPKRTELLLRAT--EKPGPAEAARQAS 304  
QY 122 WMRVSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKYKNQKTEEA 174  
DB 305 ASRAWGL-REFRSPQVLPSPD-----GRRAGV--RLAVTRLEG-VDEATRA 348

RESULT 10  
US-08-750-532-9  
Sequence 9, Application US/08750532  
Patent No. 5756339  
GENERAL INFORMATION:  
APPLICANT: MITTA, Masanori  
APPLICANT: YAMAMOTO, Katsuhiko  
APPLICANT: MORISHITA, Mio  
APPLICANT: ASADA, Kiyozo  
APPLICANT: TSUNASAWA, Susumu  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.  
STREET: 419 Seventh Street N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/750,532  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/01095  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 1994/130236  
FILING DATE: 13-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 1994/173912  
FILING DATE: 26-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: MITTA-1  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1398 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-750-532-9

Query Match 6.4%; Score 75.5; DB 1; Length 1398;  
Best Local Similarity 16.8%; Pred. No. 19;  
Matches 39; Conservative 39; Mismatches 69; Indels 85; Gaps 9;  
QY 50 DSGNKPGLPRKGLY--MANDLKLRLHHLQI----- 79  
DB 1097 DKADFAGVLTPAEGLVGEARNYTLVKHALTLEVPVPNATVIIGNTYLTDENGTVTFYA 1156  
QY 80 PIHPKDFSLVMEKGLSAMRELTAVNLEHPEMLEKASRELWVRWVRNEDITEPOSIL 139  
DB 1157 PTKLGSDEITVIVKKNFTLEKTFQITVSEPEITE-----EDINEPKLAM 1202  
QY 140 AAEEKAG--MSAE-QAOGLEKAT-----PKVKNQKLETTAAACR 177  
DB 1203 SSPEANATIVSEMESEGGVKVTVEITTINGTANETATIVVPKKAENIEVSGDHVIS 1262  
QY 178 YG-----AFGLPIITVAHVDGQTHMLFGSDRMELLA-HLGEKW 214  
DB 1263 YSIEBEGYAKYVITVKFASPTVT-----VTYTIYAGPRVSILTLEFLGYSW 1310

RESULT 11  
US-08-473-553A-3  
Sequence 3, Application US/08473553A  
Patent No. 5859338  
GENERAL INFORMATION:  
APPLICANT: Meyerowitz, Elliot M.  
APPLICANT: Clark, Steven E.  
APPLICANT: Williams, Robert W.  
TITLE OF INVENTION: Plant Clavatal Nucleic Acids,  
TITLE OF INVENTION: Transformed Plants, and Proteins  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,553A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 523 amino acids

TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-473-553A-3

Query Match 6.2%; Score 73.5; DB 2; Length 523;  
Best Local Similarity 21.8%; Pred. No. 7;  
Matches 52; Conservative 34; Mismatches 77; Indels 75; Gaps 14;  
QY 36 NQLRPSLTGIMKDSGNKPP---GLPRKGLYMANDLKLRLHHLQIPIHPKDFLSVNL 92  
DB 194 NLKHLHTFLHNNLTGHIPPELSGLVSLKSL-----DLST-----NQLTGEIPQSFNL-- 243  
QY 93 EKGSLSAMREFTAVNL-----EHPEMLEKASRELWVRWVRNEDITEPOSILAAAAE-- 143  
DB 244 --GN-----ITLINFRNNLYGQIPEAIGELPKLEVFVWENNFTIQLPANLGRNGNLI 295  
QY 144 KAGMSAEQAQGLL-----EKIATPK-----VKNLKET 171  
DB 296 KLDVSDNHLTCLIPKDLRCRGEKLEMLILSNNEFFGPIPELGGCKSLTKIRIVKNLLNGT 355  
QY 172 TENACRYGARGCLP-ITVAHV-----DQTHMLFGSDRMELLALLHLLGEKWM-GPIPPAV 222  
DB 356 VPA-----GLFNLPLVTIETLDFNFFSGELPVTMSGVDVLDQI---YLSNNWFSGEIPPAI 407

RESULT 12  
US-08-188-582-32  
Sequence 32, Application US/08188582  
Patent No. 5534410  
GENERAL INFORMATION:  
APPLICANT: Tjian, Robert  
APPLICANT: Comai, Lucio  
APPLICANT: Dynlacht, Brian D.  
APPLICANT: Hoey, Timothy  
APPLICANT: Ruppert, Siegfried  
APPLICANT: Tanese, Naoko  
APPLICANT: Wang, Edith  
APPLICANT: Weinzierl, Robert O.J.  
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFS AND METHODS OF USE  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/188,582  
FILING DATE: 28-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A.  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 869 amino acids  
TYPE: amino acid  
TOPOLOGY: linear



MOLECULE TYPE: protein  
US-08-188-582-32

Query Match 6.2%; Score 73.5; DB 1; Length 869;  
Best Local Similarity 20.5%; Pred. No. 15;  
Matches 44; Conservative 30; Mismatches 82; Indels 59; Gaps 10;

QY 32 IWNINLQRPSTLITGIMKDSGNKPPCLLPKGLYMANDKLLRHHHQIPIHFPPKDFLSVM 91  
DB 318 VMKFGKQWQPTLLQAMQVEKGATGSLP-----HLPGLAICSRSGAVCLWSPDGLRQI 373

QY 92 LEKGSLSAMRFLTAVNLEHPMLEKASRELMMRWVSRNEDITEPQSILAAAEKAGMSAEQ 151  
DB 374 -----YRDPETLVFRDSSWR--WA---DFTAHPRVLTGDRTGVMKMD 412

QY 152 AQG-----LLEKIATPKVKNLKETEACRYGA-----FG-----LPITVAHVVDG 192  
DB 413 TQGPFGCGLLFRIG-----AEASCQKGERVLLTQYLGHSPPKCLPPTLHLVCT 461

QY 193 QTHMLFGSDRMELLALLGEKWMGPPI-PAVNARL 226  
DB 462 QFSLYLVDRLPLVPM-----KWNHGLPSPLLARL 493

RESULT 13  
US-08-646-715-32  
; Sequence 32, Application US/08646715  
; Patent No. 5637886  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith  
; APPLICANT: Weinzierl, Robert O. J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS.  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/646,715  
; FILING DATE: 09-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/188,582  
; FILING DATE: 28-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-57650-2/ATT/RAO  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 869 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-646-715-32

Query Match 6.2%; Score 73.5; DB 1; Length 869;  
Best Local Similarity 20.5%; Pred. No. 15;  
Matches 44; Conservative 30; Mismatches 82; Indels 59; Gaps 10;

QY 32 IWNINLQRPSTLITGIMKDSGNKPPCLLPKGLYMANDKLLRHHHQIPIHFPPKDFLSVM 91  
DB 318 VMKFGKQWQPTLLQAMQVEKGATGSLP-----HLPGLAICSRSGAVCLWSPDGLRQI 373

QY 92 LEKGSLSAMRFLTAVNLEHPMLEKASRELMMRWVSRNEDITEPQSILAAAEKAGMSAEQ 151  
DB 374 -----YRDPETLVFRDSSWR--WA---DFTAHPRVLTGDRTGVMKMD 412

QY 152 AQG-----LLEKIATPKVKNLKETEACRYGA-----FG-----LPITVAHVVDG 192  
DB 413 TQGPFGCGLLFRIG-----AEASCQKGERVLLTQYLGHSPPKCLPPTLHLVCT 461

QY 193 QTHMLFGSDRMELLALLGEKWMGPPI-PAVNARL 226  
DB 462 QFSLYLVDRLPLVPM-----KWNHGLPSPLLARL 493

RESULT 14  
US-08-473-553A-6  
; Sequence 6, Application US/08473553A  
; Patent No. 5859338  
; GENERAL INFORMATION:  
; APPLICANT: Meyerowitz, Elliot M.  
; APPLICANT: Clark, Steven E.  
; APPLICANT: Williams, Robert W.  
; TITLE OF INVENTION: Plant Clavatal Nucleic Acids,  
; TITLE OF INVENTION: Transformed Plants, and Proteins  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/473,553A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-60886/RET/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 980 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-473-553A-6

Query Match 6.2%; Score 73.5; DB 2; Length 980;  
Best Local Similarity 21.8%; Pred. No. 18;  
Matches 52; Conservative 34; Mismatches 77; Indels 75; Gaps 10;

QY 36 NLQRPRLTGIMKDSGNKPP---GLPRKGLYMANDKLLRHHLQIPIHFPKDFLSVNL 92  
 DB 263 NLKHLHTLFHNNLTGHPPELSGLVSKSL-----DLSI-----NQTGEIPOSFINL-- 312  
 QY 93 EKGSLSAMRFTAVNL-----EHPMLEKASRELWMRYWSRNEDITEPOSILAAAE-- 143  
 DB 313 --GN-----ITLINFRNNLYGQIPEAIGELPKLEFVEVWENNFTLQLPANLGRNGNLI 364  
 QY 144 KAGMSAQOAGLL-----EKIATPK-----VKNQKXET 171  
 DB 365 KLDVSDNHLTGILPKDLCRGEKLEMLILSNFFFGPIPEELGCKCKSLTKIRIVKNLLNGT 424  
 QY 172 TEACRYGAFGLP-ITVAHV-----DQOTHMLFGSDRMELLAHLGKWM-GPIPPAV 222  
 DB 425 VPA-----GLFNLPLVTIITLTDNFFSGELPVTMSGDVLDQI--YLSNNWFSGEIPPAI 476

RESULT 15

US-08-473-553A-2  
 ; Sequence 2, Application US/08473553A  
 ; Patent No. 5859338  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Meyerowitz, Elliot M.  
 ; APPLICANT: Clark, Steven E.  
 ; APPLICANT: Williams, Robert W.  
 ; TITLE OF INVENTION: Plant Clavatal Nucleic Acids,  
 ; TITLE OF INVENTION: Transformed Plants, and Proteins  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States  
 ; ZIP: 94111-4187  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/473,553A  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 800  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Silva, Robin M.  
 ; REGISTRATION NUMBER: 38,304  
 ; REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 781-1989  
 ; TELEFAX: (415) 398-3249  
 ; TELEX: 910 277299  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 985 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-473-553A-2

Query Match 6.28; Score 73.5; DB 2; Length 985;  
 Best Local Similarity 21.88; Pred. NO. 18;  
 Matches 52; Conservative 34; Mismatches 77; Indels 75; Gaps 14;  
 QY 36 NLQRPRLTGIMKDSGNKPP---GLPRKGLYMANDKLLRHHLQIPIHFPKDFLSVNL 92  
 DB 268 NLKHLHTLFHNNLTGHPPELSGLVSKSL-----DLSI-----NQTGEIPOSFINL-- 317  
 QY 93 EKGSLSAMRFTAVNL-----EHPMLEKASRELWMRYWSRNEDITEPOSILAAAE-- 143  
 DB 318 --GN-----ITLINFRNNLYGQIPEAIGELPKLEFVEVWENNFTLQLPANLGRNGNLI 369

QY 144 KAGMSAQOAGLL-----EKIATPK-----VKNQKXET 171  
 DB 370 KLDVSDNHLTGILPKDLCRGEKLEMLILSNFFFGPIPEELGCKCKSLTKIRIVKNLLNGI 429  
 QY 172 TEACRYGAFGLP-ITVAHV-----DQOTHMLFGSDRMELLAHLGKWM-GPIPPAV 222  
 DB 430 VPA-----GLFNLPLVTIITLTDNFFSGELPVTMSGDVLDQI--YLSNNWFSGEIPPAI 481

Search completed: February 15, 2001, 15:52:22  
 Job time: 94 sec

Thu Feb 15 15:59:13 2001

us-08-978-174-1.rai

---

Page 9





A:Accession: T34201  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-225 <DUZ>  
A:Cross-references: EMBL:U41011; PIDN:AAA82289.1; CESP:D2024.7  
C:Genetics:  
A:Gene: CESP:D2024.7  
A:Introns: 51/1; 177/3

Query Match 25.2%; Score 298.5; DB 2; Length 225;  
Best Local Similarity 33.0%; Pred. No. 2.1e-18;  
Matches 74; Conservative 43; Mismatches 92; Indels 15; Gaps 7;  
QY 3 LPRTVELFYDVLSPYSLWGLFEILCRYQNIWNINLQRLPSLITGIMKDSGNKPPGL--L 59  
DB 2 PNRKVKFFEDVISYSGFEGITRHSVWKTPIQMKPFFAGVVRHTEN--PGLPLRI 59  
QY 60 PRKGLYMANDLKLRLHHLQIPIHFPKDFLSVLMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119  
DB 60 PIKEKYMHDLLFSAGYWGIFPLPKDYTNMLNTSSIVPQILVASQLRDNVLMEDVAR 119  
QY 120 ELWMRVSRNEDI--TEPQSILAAAEKAGMSAEQAQGLLEKIATPKVNKOLKETTEACRY 178  
DB 120 GLWHRFYAYGKPIFTKSQ---VAEVLRLDHLVQYDVLVMSDSAEVKNILRENTDEAIGN 176  
QY 179 GAFGLP--ITVAHVDGQT-HMFLGSDRMELLAHLGKWMGPI 218  
DB 177 GCFGAPMWHITDGH--GKVLQTVFGSDRLPQVADFLAEPPKGM 218

RESULT 3  
G83629  
hypothetical protein PA0118 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 15-Sep-2000  
A:Accession: G83629  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950  
A:Accession: G83629  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-195 <STO>  
A:Cross-references: GB:AE004450; GB:AE004091; NID:g9945943; PIDN:AAG03508.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0118

Query Match 16.7%; Score 198; DB 2; Length 195;  
Best Local Similarity 26.1%; Pred. No. 7.2e-10;  
Matches 55; Conservative 41; Mismatches 93; Indels 22; Gaps 5;  
QY 4 LPRTVELFYDVLSPYSLWGLFEILCRYQNIWNINLQRLPSLITGIMKDSGNKPPGLPRKG 63  
DB 1 MSKQIEFFDFGSGTTLATQPLRIAAHAGASIAWRPMLLGGVFKATGNHSPIEVPAK 60  
QY 64 LYMANDLKLRLHHLQIPIHFPKDFLSVLMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119  
DB 61 RYTLHDILARYAKRYGVPLAFNPAPFINTLTLM-----RGAQGYLGG-----EGFQYILK 109  
QY 120 ELWMRVSRNEDI--TEPQSILAAAEKAGMSAEQAQGLLEKIATPKVNKOLKETTEACRY 179  
DB 110 AVEALWVRQNLGKPEVVAQVLAEGFDPD---FLRVGDEQVKEGLKATTEAVRRG 166  
QY 180 AFGLPITVAHVDGQTHMFLGSDRMELLAHL 210  
DB 167 VEGAPSFV---GDQLFFGGDRLDFAEVL 193

## RESULT 4

S72164  
2-hydroxychromene-2-carboxylate isomerase homolog - Rhizobium leguminosarum bv. C:Species: Rhizobium leguminosarum bv. viciae  
C:Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 29-Sep-1999  
C:Accession: S72164  
R:Brito, B.; Palacios, J.M.; Ruiz-Argüeso, T.; Imperial, J.  
Biochim. Biophys. Acta 1308, 7-11, 1996  
A:Title: Identification of a gene for a chemoreceptor of the methyl-accepting A:Reference number: S72162; MUID:96328256  
A:Accession: S72164  
A:Molecule type: DNA  
A:Residues: 1-194 <BRI>  
A:Cross-references: EMBL:U23040; NID:g780654; PIDN:AAC44312.1; PID:g780657  
A:Experimental source: strain UPM791  
C:Genetics:  
A:Gene: plasmid  
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 13.9%; Score 164.5; DB 2; Length 194;  
Best Local Similarity 25.2%; Pred. No. 5.4e-07;  
Matches 54; Conservative 34; Mismatches 91; Indels 35; Gaps 8;  
QY 6 RTVELFYDVLSPYSLWGLFEILCRYQNIWNINLQRLPSLITGIMKDSGNKPPGL-LPRKGL 64  
DB 2 RTLDFYDYSYSLALSQVRK---MDVEIAFHPLLEIGLDMKQVGNVPTSITCAPKGR 57  
QY 65 YMANDLKLRLHHLQIPIHFPKDFLSV-----MLEKSGLSAMRFLTAVNLEHPEMLEK 116  
DB 58 YVMTDIQRWAVHYGVSLNWHPPOLLEIDASRLRLATLVAGOLGAMP--TAV----- 105  
QY 117 ASRELWVRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVNKOLKETTEAAC 176  
DB 106 --EAFINAIWASAPLATAAEVAALGAAGLDAEE---LAERMDPEAAQDLLEATANAV 160  
QY 177 RYGAFLPITVAHVDGQTHMFLGSDRMELL-AHL 209  
DB 161 SRGVEGAPTFLV---GDEMFEGNDRLHFMOGHL 190

RESULT 5  
T31286  
2-nitrotoluene dioxygenase (EC 1.14.-.-) Rieske iron-sulfur component - Sphingomonas C:Species: Sphingomonas aromaticivorans  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000  
C:Accession: T31286  
R:Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, submitted to the EMBL Data Library, July 1998  
A:Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas A:Reference number: Z20992  
A:Accession: T31286  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-197 <ROM>  
A:Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378427; PIDN:AAD04010.1  
C:Genetics:  
A:Gene: nahD  
A:Gene: plasmid pN1  
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase  
C:Keywords: oxidoreductase

Query Match 11.8%; Score 140; DB 2; Length 197;  
Best Local Similarity 23.1%; Pred. No. 6.9e-05;  
Matches 50; Conservative 40; Mismatches 94; Indels 32; Gaps 7;  
QY 4 LPRTVELFYDVLSPYSLWGLFEILCRYQNIWNINLQRLPSLITGIMKDSGNKPPG---LLP 60  
DB 1 MTRTHFYDFISPSYLAQLKLEIARAAAGCTVDYWDIPIDPEAKIAAGNYSNREVL 6C  
QY 61 RKGLYMANDLKLRLHHLQIPIHFPKDF-----LSVMLEKSGLSAMRFLTAVNLEHPEML 114

123 MRYWSRNEITEPOSILAA--AEKAGSAEQAGGLEKIATPKVKNQKLETTAAACRYGA 180  
111 NAVM--GEGIAPDLESPLVSEKLGWDSAEHEFLSSNAATE--RYDSQTHAAIERKV 165  
181 FGLPITVAHVGDQTHMLFGSDRMELLALLG 211  
166 FGVP-TWFLGO---EMWGNDRFLMESANG 192

RESULT 8  
H72425  
ABC transporter, ATP-binding protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: H72425  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.;  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richards  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genomic  
A:Reference number: A72200; MUID:99287316  
A:Accession: H72425  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-581 <ARN>  
A:Cross-references: GB:AE001691; GB:AE000512; NID:g4980517; PIDN:AAD35137.1; PID:  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0043  
C:Superfamily: Escherichia coli ABC transporter mdIA; ATP-binding cassette hcmJ

Query Match 7.7%; Score 91; DB 2; Length 581;  
Best Local Similarity 19.0%; Pred. No. 4.5;  
Matches 35; Conservative 44; Mismatches 67; Indels 38; Gaps 5;  
QY 28 RYONIW-----NINQLRPSLITGIMKDSGNKPPGL-----LPRKGLYMAN 68  
DB 343 REENWFSDGKNWVLKIDINLQFGKLYAIVGETGGKSTLMSLNGILYIPKGNIFD 402

QY 69 DLKLLRHHLIQIP---IHFPKDFL-----SYMLEKSGLSAMRFLTAVNLEHP-EMLE 115  
DB 403 EPLLEYNKLVKQTAANVDVLLFSGTILNIRLFDSEIPEERVLEAKRVAHLDIIE 462

QY 116 KASRELWVRWSRNEITEPOSILAAAEKAGM-----SAGQAGLLEKIATPKVKNQK 169  
DB 463 RLPGGVYIEVERGTLSAGERQLIALARAVLFDKIFILDEATSNVDVITETKIOEAL 522

QY 170 EITE 173  
DB 523 ELSK 526

RESULT 9  
C69127  
flavoprotein A homolog (II) - Methanobacterium thermoautotrophicum (strain Delta H  
C:Species: Methanobacterium thermoautotrophicum  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
C:Accession: C69127  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldred,  
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiv  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H  
A:Reference number: A69000; MUID:98037514  
A:Accession: C69127  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-410 <MB>  
A:Cross-references: GB:AE000809; GB:AE000666; NID:g2621265; PIDN:AAB84726.1; PID:  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH220

61 -KIKVMKADLERWAERYGVPLTFPASFACADWNCVLFAREHGKAEAFV-----109  
115 EKASRELWVRWSRNEITEPOSILAAAEKAGSAEQAGGLEKIATPKVKNQKLETTAA 174  
110 -----DAIRRWGIGIDPDGRNEIAACAIAAGL---DPAALIAFVESPAGQNEYRKARSO 161  
175 ACRYGAFGLPITVAHVGDQTHMLFGSDRMELLALLH 210  
162 AIQGVVYGAFL--MFVDDQ--IFWGNDRDLFLAEYL 193

RESULT 6  
S17164  
glutathione transferase (EC 2.5.1.18) 13 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 21-Nov-1993 #sequence\_revision 13-Mar-1997 #text\_change 13-Mar-1997  
C:Accession: S17164  
R:Haris, J.M.; Meyer, D.J.; Coles, B.; Ketterer, B.  
Biochem. J. 278, 137-141, 1991  
A:Title: A novel glutathione transferase (13-13) isolated from the matrix of rat liver  
A:Reference number: S17164; MUID:91354194  
A:Accession: S17164  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-33 <HAR>  
C:Keywords: transferase

Query Match 10.5%; Score 125; DB 2; Length 33;  
Best Local Similarity 71.9%; Pred. No. 0.00013;  
Matches 23; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
QY 3 PLPRTVELFYDVLSPYSWLGFEILCRYQNIWN 34  
DB 2 PAPRVLELYDVLSPYSXGLGFEVLRYQHLXN 33

RESULT 7  
C55552  
2-hydroxychromene-2-carboxylate isomerase - Pseudomonas putida plasmid NAH7  
C:Species: Pseudomonas putida  
C>Date: 08-Sep-1995 #sequence\_revision 08-Sep-1995 #text\_change 29-Sep-1999  
C:Accession: C55552  
R:Eaton, R.W. 1994  
J. Bacteriol. 176, 7757-7762, 1994  
A:Title: Organization and evolution of naphthalene catabolic pathways: sequence of the  
from the NAH7 plasmid.  
A:Reference number: A55552; MUID:95095951  
A:Accession: C55552  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-203 <EAT>  
A:Cross-references: GB:U09057; NID:g483790; PIDN:AAA66358.1; PID:g483793  
C:Genetics:  
A:Gene: nahD  
A:Genome: plasmid  
A:Start codon: GTG  
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 8.0%; Score 95; DB 2; Length 203;  
Best Local Similarity 24.2%; Pred. No. 0.53;  
Matches 51; Conservative 35; Mismatches 97; Indels 28; Gaps 8;  
QY 8 VELYDVLSPYSWLGFEILCRYQNIWNINLQRLPSLITGIMKDSGNKPPGL--LPRKGLY 65  
DB 3 VDFYDFLSPFSYLANQRLSKLAQDYGLTIRYNDAIDARVTAIGNVGSNRLDKVKLDY 62

QY 66 MANDLKLRLHHLQIPHPKDFLSVMLEK-----SLSAMRFLTAVNLEHPMELEKASREWL 122  
DB 63 LKVDLQRAQLYGLIPFPANYSRRMNGFYSGAEAAQAAAYNV-----VF 110

A:Start codon: TTG  
C:Superfamily: Methanobacterium flavoprotein A  
C:Keywords: flavoprotein

Query Match 7.6%; Score 90.5; DB 2; Length 410;  
Best Local Similarity 25.1%; Pred. No. 3.2;  
Matches 65; Conservative 31; Mismatches 82; Indels 81; Gaps 16;  
QY 13 DVLSPSWLGFEILC-----RYQNIWNINLQRLPSLTIGIMKDSGNK----- 54  
DB 99 EVLKRY---GSEICTAKAAGLKHYSIPQDTPMQ---TVKTGSTDLGGKTLTFLEAP 152  
QY 55 -----PPG---LLPRKGLYMANDLKLRLHLLQIPHPKDFLSVMLEKGSLSAMRF---L 103  
DB 153 MLHWPDSPMTLLEEGILFSND--AFGQHLCSKRFKDVPEAVLMD---AAKMFYANLL 207  
QY 104 TAVNLEHPEMLERAS--RELWM-----RVWSRNEDEITPEQSILAAAEKAGMSA 149  
DB 208 TPLS---PLVRFSEVKEGLLEKIGMIAPSHGQIW-----TEPLKIIAA-----YT 252  
QY 150 EQAQGLLEKIATPKVNQLKETTAAACRYGAFGLPITVAHVGDQTHMLFGSDRMELLAHL 209  
DB 253 DWATGCRDKAT--IYDTMHSYEMLAHAAEGL--MAADVDSMHFLHEDERSEIVKNI 309  
QY 210 LGEK-----WMGPPIP 219  
DB 310 LESKAVFIGSPMTFNGPFP 328

RESULT 10  
SNECPI  
N:pitilysin (EC 3.4.24.55) precursor [validated] - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 31-Mar-1993 #sequence\_revision 31-Oct-1997 #text\_change 18-Feb-2000  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: F65064  
A:Molecule type: DNA  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Residues: 1-962 <BLAT>  
A:Cross-references: GB:A5000365; GB:U00096; NID:G2367163; PIDN:AAC75860.1; PID:G2367164;  
A:Experimental source: Strain K-12, substrain MG1655  
R:Claverie-Martin, F.; Diaz-Torres, M.R.; Kushner, S.R.  
Gene 54, 185-195, 1987  
A:Title: Analysis of the regulatory region of the protease III (ptr) gene of Escherichia coli.  
A:Reference number: A29093; MUID:88005781  
A:Accession: A29093  
A:Molecule type: DNA  
A:Residues: 1-276; 'HYHSLR', 283, 'W', 285-296 <CLA>  
A:Cross-references: GB:M17095; NID:G147390; PIDN:AAA24436.1; PID:G147391  
A:Experimental source: strain K12  
R:Finch, P.W.; Wilson, R.E.; Brown, K.; Hickson, I.D.; Emmerson, P.T.  
Nucleic Acids Res. 14, 7695-7703, 1986  
A:Title: Complete nucleotide sequence of the Escherichia coli ptr gene encoding protease III.  
A:Reference number: A25765; MUID:87040734  
A:Accession: A25765  
A:Molecule type: DNA  
A:Residues: 1-962 <FTN>  
A:Cross-references: GB:X06227; NID:G42560; PIDN:CAA29576.1; PID:G42561  
R:Becker, A.B.; Roth, R.A.  
Proc. Natl. Acad. Sci. U.S.A. 89, 3835-3839, 1992  
A:Title: An unusual active site identified in a family of zinc metalloendopeptidases.  
A:Reference number: A38854; MUID:92237263  
C:Contents: annotation; active site  
C:Genetics:  
A:Gene: ptr

A:Map position: 61

C:Function:

A:Description: endopeptidase degrades small peptides [validated; MUID:92237263]  
A:Pathway: protein degradation  
C:Superfamily: insulysin  
C:Keywords: hydrolase; metalloproteinase; monomer; periplasmic space; protein de  
F:1-23/Domain: signal sequence #status predicted <Sig>  
F:24-962/Product: pitrilysin #status experimental <MAT>  
F:88,92/Binding site: zinc (His) #status experimental  
F:91/Active site: Glu #status experimental

Query Match 7.6%; Score 89.5; DB 1; Length 962;  
Best Local Similarity 26.3%; Pred. No. 12;  
Matches 30; Conservative 13; Mismatches 48; Indels 23; Gaps 3;  
QY 126 WSRNEDITEPQISILAAAEKAGMSAEQA-----QGLEKIATPKVNQ 167  
DB 731 WCRNKDVVYDKQSVIFERKAGNSTDSALAAVFVPTGYDEYTSAYSSLLGQIVQPFYQ 790  
QY 168 LKETTEACRYGAFGLPITVAHVGDQTHMLFGSDRMELLAHLGKMGPIPPA 221  
DB 791 LR--TEEQGLYAVFAFPMVSVGQMGFLQSDNQ----PSFLWERYKAFFETA 839

RESULT 11  
I49343  
A:probable isomerase doxJ - Pseudomonas sp. (strain Cl8)  
C:Species: Pseudomonas sp.  
C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 18-Sep-1998  
R:Denome, S.A.; Stanley, D.C.; Olson, E.S.; Young, K.D.  
J. Bacteriol. 175, 6890-6901, 1993  
A:Title: Metabolism of dibenzothiophene and naphthalene in Pseudomonas strains:  
A:Reference number: A49343; MUID:94042852  
A:Accession: I49343  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-199 <DEN>  
A:Cross-references: GB:M60405  
C:Genetics:  
A:Gene: doxJ  
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 7.5%; Score 89; DB 2; Length 199;  
Best Local Similarity 24.2%; Pred. No. 1.7;  
Matches 51; Conservative 37; Mismatches 95; Indels 28; Gaps 9;

QY 8 VELYDVLSYSPSWLGFEILCRYQNIWNINLQRLPSLTIGIMKDSGNKPPGL--LPRKGLY 65  
DB 3 VDFYDFELSPSYLANHRLSKLAQDYGFISIRYAIOLARVKIAIGNVGSPNRDLIVKLDY 62  
QY 66 MANDKLRLRHLLQIPHPKDFLSVMLEKGSLS--SAMRELTAVNLEHPEMLEKASRELW 122  
DB 63 LKVDLQRAEYIEPLVFPYANYSRRMTGLYSGAMAAQTGAYVNV-----VF 110  
QY 123 MRYWSRNEDEITPEQSILAA--AEKAGMSAEQAQGLEKIATPKVNQLKETTEAACRYGA 180  
DB 111 NAVW--CGDIAPDLESPLVSEKLGWDSAFE---DFISSDAATERYDEQTHAAIERKV 155  
QY 181 FGLPITVAHVGDQTHMLFGSDRMELLAHLG 211  
DB 166 FGVP-TMFLG---EMWGNDRFLFLENVAVG 192

RESULT 12  
T51712

A:threonine dehydratase/deaminase [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
C:Accession: T51712  
R:Mourad, G.S.; Emerick, R.M.; Marion, A.L.; Smith, A.M.



submitted to the EMBL Data Library, October 1998  
A:Reference number: Z25433  
A:Accession: J51712  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-592 <MOU>  
A:Cross-references: EMBL:AF096281; PIDN:AAC97936.1  
C:Genetics:  
A:Gene: OMR1

Query Match 7.5%; Score 89; DB 2; Length 592;  
Best Local Similarity 22.8%; Pred. No. 6.9;  
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
QY 55 PPGI-LPRKGLYMANDLKLRLHLLQIPHFPRKDFLSVM-----LEKGSLS-AMRELT-- 104  
Db 52 PKKLPPLPR-----LKVSNSLQ-----YPAGILGAVPERTNEAENGSIAMAEYLTNI 99  
QY 105 -----AVNLEHP-EMLEKASRELMMRWVSRNEDITE-----PQSILAA 141  
Db 100 LSTKVYDIATIESPLQAKLKLKRLGVRMYLKRDLQPVFSLKRGAYNNMMVKLPADOLAK 159  
QY 142 AERKAGSAEQAG-----LLEKIATPKVKNOLKETTEAACRYGAFGLPITVA 188  
Db 160 GVICSSAGNHAOGVALSASKLGCTAVIVNPVTTPKIKQAVENL----- 203  
QY 189 HVDGOTHLFGSDRMELLAHL-----LGEKWMGPIPP 220  
Db 204 --CATVVLFGSDVDOQAQAHAKIRAEGLTEIPP 235

Query Match 7.3%; Score 87; DB 2; Length 926;  
Best Local Similarity 26.4%; Pred. No. 18;  
Matches 46; Conservative 27; Mismatches 59; Indels 42; Gaps 11;  
QY 84 PKDFLSVMLEKGSLSAMRFLTAVN-LEHPEMLEKASRELMMRWVSRNE----DITEPQSI 138  
Db 357 PEDWLC-----DGTGYDFMNQVSLQHDPRGRPLRELWQVSRPEAFDELVEYQARQL 411  
QY 139 LAAAEKAGSAEQAGLLEKIATPKVKNOLKETTEAACRYGAFGL-----PI--TVAHVDG 192  
Db 412 VLAGSLAGDLENLAQGLL--RVARADLAS--RDLTLGIRRALFOLLARFPYRTYAGACG 468  
QY 193 QT-----HMLFGSDRMELLAHLGKWMG-----PIPPAVNARL 226  
Db 469 RSVQDREVFYAAEAAREDDLEADR-AVLDHL--ERWLGGQPLRELPPGLRRL 519

Query Match 7.5%; Score 88.5; DB 2; Length 613;  
Best Local Similarity 23.6%; Pred. No. 8;  
Matches 39; Conservative 26; Mismatches 41; Indels 59; Gaps 9;  
QY 6 RTVELFYDLSPYSWLGFEILCRQYIWNIT-----NLQRLPSLTIGMKDSGNKPPGLLP 60  
Db 333 RAIRLFEKPLDP-----QSIYQLIEISRLNQLPPELDLMLK-AGEKP--VEP 377  
QY 61 RGLYMANDLKLRLHLLQIPHFPRKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120  
Db 378 EQ-----ELEIPEV-PED-----ISEVDLHDHDFLQFN----- 403  
QY 121 LMMRWVSRNEDITEPQSILAAAEKAGSAEQAGLL--EKIATPK 163  
Db 404 ---KMLSKNGYKAPGHAVALPEGLSVEDILNLLGMSAANPK 445

RESULT 13  
A35296  
secretogranin II precursor - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 17-Aug-1990 #sequence\_revision 17-Aug-1990 #text\_change 24-Sep-1999  
C:Accession: A35296  
R:Fischer-Colbrie, R.; Gutierrez, J.; Hsu, C.M.; Iacangelo, A.; Eiden, L.E.  
J. Biol. Chem. 265, 9208-9213, 1990  
A:Title: Sequence analysis, tissue distribution and regulation by cell depolarization, a  
A:Reference number: A35296; MUID:90264409  
A:Accession: A35296  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-613 <FIS>  
A:Cross-references: GB:J05468; NID:g163709; PIDN:AAA30760.1; PID:g163710  
C:Superfamily: secretogranin II

Query Match 7.1%; Score 84.5; DB 2; Length 316;  
Best Local Similarity 23.3%; Pred. No. 7.4;  
Matches 41; Conservative 30; Mismatches 60; Indels 45; Gaps 8;  
QY 13 DVLSPYSWLGFEILCRQYIWNITLQLRPSLTIGMKDSGNKPPGLLPKGLYMANDLKL 72  
Db 80 DVCDPYE--EYNVFRFQAVYGI---VPSIL-----RAHKVPIIVGCTGLYLD---AV 124  
QY 73 LRHHLOIPIHFPPKDFLSVMLEKGSLSAM-----RPLTAVNL- 108  
Db 125 LRQYALVPVE--RNQALRASLRGASLSHMRVYFSLKDSHAVHNKTDLEDPARLMRAIEIA 183  
QY 109 -----EHPMLEKASRELMMRWVSRNEDITEPQSILAAAEKAGSAEQAGLLEKIA 160  
Db 184 VFHATHPPELLQQA--RETRPMRAKVYGIQYPRSMRLARIRARLEQIRINGLIEVA 238

Query Match 7.1%; Score 87; DB 2; Length 926;  
Best Local Similarity 26.4%; Pred. No. 18;  
Matches 46; Conservative 27; Mismatches 59; Indels 42; Gaps 11;  
QY 84 PKDFLSVMLEKGSLSAMRFLTAVN-LEHPEMLEKASRELMMRWVSRNE----DITEPQSI 138  
Db 357 PEDWLC-----DGTGYDFMNQVSLQHDPRGRPLRELWQVSRPEAFDELVEYQARQL 411  
QY 139 LAAAEKAGSAEQAGLLEKIATPKVKNOLKETTEAACRYGAFGL-----PI--TVAHVDG 192  
Db 412 VLAGSLAGDLENLAQGLL--RVARADLAS--RDLTLGIRRALFOLLARFPYRTYAGACG 468  
QY 193 QT-----HMLFGSDRMELLAHLGKWMG-----PIPPAVNARL 226  
Db 469 RSVQDREVFYAAEAAREDDLEADR-AVLDHL--ERWLGGQPLRELPPGLRRL 519

Query Match 7.1%; Score 84.5; DB 2; Length 316;  
Best Local Similarity 23.3%; Pred. No. 7.4;  
Matches 41; Conservative 30; Mismatches 60; Indels 45; Gaps 8;  
QY 13 DVLSPYSWLGFEILCRQYIWNITLQLRPSLTIGMKDSGNKPPGLLPKGLYMANDLKL 72  
Db 80 DVCDPYE--EYNVFRFQAVYGI---VPSIL-----RAHKVPIIVGCTGLYLD---AV 124  
QY 73 LRHHLOIPIHFPPKDFLSVMLEKGSLSAM-----RPLTAVNL- 108  
Db 125 LRQYALVPVE--RNQALRASLRGASLSHMRVYFSLKDSHAVHNKTDLEDPARLMRAIEIA 183  
QY 109 -----EHPMLEKASRELMMRWVSRNEDITEPQSILAAAEKAGSAEQAGLLEKIA 160  
Db 184 VFHATHPPELLQQA--RETRPMRAKVYGIQYPRSMRLARIRARLEQIRINGLIEVA 238

Query Match 7.1%; Score 84.5; DB 2; Length 316;  
Best Local Similarity 23.3%; Pred. No. 7.4;  
Matches 41; Conservative 30; Mismatches 60; Indels 45; Gaps 8;  
QY 13 DVLSPYSWLGFEILCRQYIWNITLQLRPSLTIGMKDSGNKPPGLLPKGLYMANDLKL 72  
Db 80 DVCDPYE--EYNVFRFQAVYGI---VPSIL-----RAHKVPIIVGCTGLYLD---AV 124  
QY 73 LRHHLOIPIHFPPKDFLSVMLEKGSLSAM-----RPLTAVNL- 108  
Db 125 LRQYALVPVE--RNQALRASLRGASLSHMRVYFSLKDSHAVHNKTDLEDPARLMRAIEIA 183  
QY 109 -----EHPMLEKASRELMMRWVSRNEDITEPQSILAAAEKAGSAEQAGLLEKIA 160  
Db 184 VFHATHPPELLQQA--RETRPMRAKVYGIQYPRSMRLARIRARLEQIRINGLIEVA 238

RESULT 14  
E83375  
probable glycosyl hydrolase PA2162 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 15-Sep-2000  
C:Accession: E83375

Thu Feb 15 15:59:14 2001

us-08-978-174-1.rpr

Page 6

Search completed: February 15, 2001, 15:52:42  
Job time: 114 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2001, 15:50:48 ; Search time 9.96 Seconds  
(without alignments)  
732.778 Million cell updates/sec

Title: US-08-978-174-1

Perfect score: 1185

Sequence: 1 MGPLEPRTVEFDVLSYWSW.....AHLGKWMGPIPPAVNARL 226

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	869	73.3	225	1	GTK1_RAT
2	328.5	27.7	226	1	YS21_CAEEL
3	298.5	25.2	225	1	YIS7_CAEEL
4	95	8.0	203	1	NAHD_PSEPU
5	89.5	7.6	962	1	PTRA_ECOLI
6	89	7.5	212	1	NAHD_PSESP
7	88.5	7.5	613	1	SG2_BOVIN
8	84.5	7.1	316	1	MIAA_TREPA
9	84.5	7.1	971	1	YMBP_YEAST
10	83.5	7.0	488	1	CATA_LISSE
11	83.5	7.0	700	1	NCD_DROME
12	81.5	6.9	619	1	SG2_RAT
13	81.5	6.9	1061	1	DPOL_ADE12
14	81	6.8	274	1	YG73_SYNY3
15	80.5	6.8	339	1	PURK_BRUME
16	80.5	6.8	705	1	CC5_YEAST
17	80.5	6.8	1073	1	HSEB_PIG
18	80	6.8	540	1	TOP1_AQUAE
19	80	6.8	1162	1	BXEN_CLOAB
20	79.5	6.7	547	1	CATX_BACSU
21	79.5	6.7	1073	1	HSEB_HUMAN
22	78.5	6.6	383	1	PNAI_RICPR
23	78	6.6	1162	1	BXEN_CLOAB
24	77.5	6.5	634	1	YC36_METUA
25	77	6.5	494	1	ADRO_RAT
26	76.5	6.5	962	1	YBX7_SCHPO
27	76.5	6.5	1056	1	DPOL_ADE02
28	76.5	6.5	1193	1	DPOL_ADE01
29	76	6.4	405	1	PK3_HUMAN
30	76	6.4	645	1	NODQ_RHIS3
31	76	6.4	2208	1	POLN_MANCV
32	75.5	6.4	744	1	GYRA_AQUAE
33	75.5	6.4	887	1	GLND_KLEPN

Query Match 73.3% ; Score 869; DB 1; Length 225;

## ALIGNMENTS

RESULT 1

ID	GTK1_RAT	STANDARD;	PRT;	225 AA.
AC	P24473; O09034;			
DT	01-MAR-1992 (Rel. 21, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	GLUTATHIONE S-TRANSFERASE, MITOCHONDRIAL (EC 2.5.1.18) (GST 13-13)			
DE	(GLUTATHIONE S-TRANSFERASE SUBUNIT 13) (GST CLASS-KAPPA).			
GN	GTK1 OR GSTK1-1.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER;			
RX	MEDLINE=97079244; PubMed=8920976;			
RA	Pemble S.E., Wardle A.F., Taylor J.B.;			
RT	"Glutathione S-transferase class kappa: characterization by the			
RT	cloning of rat mitochondrial GST and identification of a human			
RT	homologue."			
RL	Biochem. J. 319:749-754(1996).			
RN	[2]			
RP	SEQUENCE OF 1-33.			
RX	MEDLINE=91354194; PubMed=1883325;			
RA	Harris M.J., Meyer D.J., Coles B., Ketterer B.;			
RT	"A novel glutathione transferase (13-13) isolated from the matrix of			
RT	rat liver mitochondria having structural similarity to class theta			
RT	enzymes."			
RL	Biochem. J. 278:137-141(1991).			
CC	FUNCTION: MIGHT CONFER PROTECTION AGAINST GENOTOXIC AND CYTOTOXIC			
CC	ELECTROPHILES IN THE MITOCHONDRIAL COMPARTMENT.			
CC	FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER			
CC	OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.			
CC	CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.			
CC	SUBUNIT: HOMODIMER.			
CC	SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.			
CC	SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announc/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL: S83436; AAB50831.1; ..			
DR	PIR: S17164; S17164.			
KW	Transferase; Multigene family; Mitochondrion.			
FT	INIT MET 0 0			
FT	CONFLICT 1 1 G -> C (IN REF. 2).			
SQ	SEQUENCE 225 AA; 25362 MW; FC895B730655E0C9 CRC64;			

34	75.5	6.4	1056	1	DPOL_ADE05	P04495 human adeno
35	75	6.3	231	1	LOLD_NEIMA	P57030 neisseria m
36	75	6.3	393	1	HEMX_ECOLI	P09127 escherichia
37	75	6.3	593	1	KPYA_TOBAC	O40545 nicotiana t
38	75	6.3	631	1	OE66_NPVLS	P89518 leucania se
39	75	6.3	686	1	MEPD_PIG	P47788 sus scrofa
40	75	6.3	925	1	PIPI_YEAST	P40020 saccharomyc
41	75	6.3	984	1	SECA_AQUAE	O67718 aquifex aeo
42	74.5	6.3	285	1	LAFT_VIBPA	Q03477 vibrio para
43	74.5	6.3	402	1	G3PB_ARATH	P25857 arabidopsis
44	74.5	6.3	545	1	FGR_FSVGR	P00544 feline sarc
45	74.5	6.3	888	1	LONI_ARATH	O64948 arabidopsis

Best Local Similarity 69.3%; Pred. No. 1.5e-68;  
Matches 156; Conservative 38; Mismatches 31; Indels 0; Gaps 0;

QY 2 GPLPRTVELFDVLSYSLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLPR 61  
pb 1 GPAPRVLEFDVLSYSLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLPR 60  
QY 62 KGLYMANDLKLRLHLLQIPHFPPKDFLSVLMLEKGLSAMRFLTAVNLEHPEMLEKASREL 121  
Db 61 KGQYLKEIPLKQLFQVPMSPKDFEHEVKKGTVMARFELTAVSMEQPEMLEKVSREL 120  
QY 122 WMRVWSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETAAACRYGAF 181  
Db 121 WMRVWSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETAAACRYGAF 180  
QY 182 GLPITVAHVGGTHMLFGSDRMELLALHLLGKMKGPPIPPAVNARL 226  
Db 181 GLPITVAHVGGTHMLFGSDRMELLALHLLGKMKGPPIPPAVNARL 225

RESULT 2  
YS21 CAEEL STANDARD; PRT; 226 AA.  
AC Q09652;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE HYPOTHETICAL 26.5 KDA PROTEIN ZK1320.1 IN CHROMOSOME II.  
GN ZK1320.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Berks M.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: Z46934; CAA87039.1;  
DR WORMPEP: ZK1320.1; CE01698.  
KW Hypothetical protein.  
SQ SEQUENCE 226 AA; 26482 MW; DBA39A5994300164 CRC64;

Query Match 27.7%; Score 328.5; DB 1; Length 226;  
Best Local Similarity 33.0%; Pred. No. 1.3e-21;  
Matches 73; Conservative 47; Mismatches 94; Indels 7; Gaps 5;

QY 1 MGPLPRTVELFDVLSYSLGFEILCRYQNIW-NINLQRLPSLITGIMKDSGNKPPGL 59  
Db 1 MPKLPRI-IDFVFDVLSYSLGFEILCRYQNIW-NINLQRLPSLITGIMKDSGNKPPGL 59  
QY 60 PRKGLYMANDLKLRLHLLQIPHFPPKDFLSVLMLEKGLSAMRFLTAVNLEHPEMLEKASR 119  
Db 60 PARSIMMTDLKRTAKFDWIDPLTPPFLMEWIKKRTTGAMKVLLVLEQDKELMLRAAR 119  
QY 120 ELWVRVWSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETAAACRYG 179  
Db 120 EMVRLWSRSKIEFDQDFEVLKAVGV--KNPEQIVEKSKDEKVIKILMENTNGVDLM 177  
QY 180 AFGLP-ITVAHVGGTHMLFGSDRMELLALHLLGKMKGPPI 219  
Db 178 AYGAPEWNVHTEDGSEHFFGSDRFLIADLLQOP--QPLP 216

RESULT 3  
YYS7 CAEEL STANDARD; PRT; 225 AA.  
AC Q18973;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHETICAL 26.0 KDA PROTEIN D2024.7 IN CHROMOSOME IV.  
GN D2024.7.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Du Z.; Gattung S.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: U41011; AAA82289.1;  
DR WORMPEP: D2024.7; CE04296.  
KW Hypothetical protein.  
SQ SEQUENCE 225 AA; 2300A3D6762B7232 CRC64;

Query Match 25.2%; Score 298.5; DB 1; Length 225;  
Best Local Similarity 33.0%; Pred. No. 5.2e-19;  
Matches 74; Conservative 43; Mismatches 92; Indels 15; Gaps 7;

QY 3 PLPRTVELFDVLSYSLGFEILCRYQNIW-NINLQRLPSLITGIMKDSGNKPPGL---L 59  
Db 2 PNKRVKFEFDVLSYSLGFEILCRYQNIW-NINLQRLPSLITGIMKDSGNKPPGL 59  
QY 60 PRKGLYMANDLKLRLHLLQIPHFPPKDFLSVLMLEKGLSAMRFLTAVNLEHPEMLEKASR 119  
Db 60 PIKEYMHKDLFLSAQWGPFPPLPKDYTNMMLNTSSIVPQRLVASQLRDNVLMEDVAR 119  
QY 120 ELWVRVWSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETAAACRY 178  
Db 120 GLWHRFYAYGKPIFTKSQ---VAEVLRLDHVRKVDVDELVMMSDSAEVKNILRENTDEATGN 176  
QY 179 GAFGLP---ITVAHVGGTHMLFGSDRMELLALHLLGKMKGPPI 218  
Db 177 GCFGAPWMTDGH--GKVLQTVFGSDRLPOVADFLAEFPKGP 218

RESULT 4  
NAHD\_PSEPU STANDARD; PRT; 203 AA.  
AC Q51948;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE (HCCA ISOMERASE).  
GN NAHD.  
OS Pseudomonas putida.  
OC Plasmid NAH7.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-G7 / ATCC 17485;  
RX MEDLINE-95095951; PubMed-8002605;  
RA Eaton R.W.;



DR PROSITE; PS00143; INSULINASE; 1.  
 KW Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 962  
 FT METAL 88 88  
 FT ACT\_SITE 91 91  
 FT METAL 92 92  
 FT METAL 169 169  
 FT METAL 88 88  
 FT MUTAGEN 91 91  
 FT MUTAGEN 92 92  
 FT MUTAGEN 162 162  
 FT MUTAGEN 169 169  
 FT MUTAGEN 204 204  
 FT CONFLICT 277 284  
 SQ SEQUENCE 962 AA; 107708 MW; 0558682C2F1A0540 CRC64;

Query Match 7.6%; Score 89.5; DB 1; Length 962;  
 Best Local Similarity 26.3%; Pred. No. 4.4;  
 Matches 30; Conservative 13; Mismatches 48; Indels 23; Gaps 3;

QY 126 WSRNEDITEPQSILAAAEKAGMSAEQA-----QGLLEKIATPKYKNQ 167  
 DB 731 WCRNKDVVDKOSVIFEKAGNSTDSALAAVFTGYDEYTSAYSSLLGQIVQPFYQ 790  
 QY 168 LKETTEAACRYGAFGLPITVAHVGDQTHMFLGSDRMELLALLHLLGKWMGP 221  
 DB 791 LR--TEEQLYGAVFAPMSVGRQWGMGFLLOSNDKQ---PSFLWERYKAFPTA 839

RESULT 6  
 NAHD\_PSESP STANDARD; PRT; 212 AA.  
 ID NAHD\_PSESP  
 AC Q52462;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE (HCCA ISOMERASE).  
 GN DOXJ.  
 OS Pseudomonas sp. (strain C18).  
 OC Plasmid.  
 OC Bacteria; Proteobacteria.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94042852; PubMed-8226631;  
 RA Denome S.A., Stanley D.C., Olson E.S., Young K.D.;  
 RT "Metabolism of dibenzothioephene and naphthalene in Pseudomonas strains: complete DNA sequence of an upper naphthalene catabolic pathway";  
 RT J. Bacteriol. 175:6890-6901(1993).  
 CC -1- FUNCTION: CATALYZES THE ISOMERIZATION OF 2-HYDROXYCHROMENE-2-CARBOXYLATE (HCCA) TO TRANS-O-HYDROXYBENZYLIDENEPYRUVATE (THBPA). THE OPTIMUM PH FOR THE ENZYME IS 10. THE REACTION IS REVERSIBLE.  
 CC -1- PATHWAY: UPPER NAPHTHALENE CATABOLIC PATHWAY WHICH INVOLVES CONVERSION OF NAPHTHALENE TO SALICYLATE, AND CATABOLISM OF DIBENZOTHIOEPHENE (DBT) AND PHENANTHRENE. CONVERTS PHENANTHRENE TO 1-HYDROXY-2-NAPHTHOIC ACID AND THE METABOLISM OF DBT IS LIMITED TO OXIDATION OF THE AROMATIC RING.  
 CC -1- MISCELLANEOUS: DOXH AND DOXJ ENCODE DIFFERENT ENZYMES THAT MAY HAVE INTERCHANGEABLE FUNCTIONS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC EMBL; M60405; AAA16133.1;  
 DR Isomerase; Plasmid; Aromatic hydrocarbons catabolism.  
 KW Isomerase; Plasmid; Aromatic hydrocarbons catabolism.  
 SQ SEQUENCE 212 AA; 24039 MW; 5EF96A619913DB4F CRC64;

Query Match 7.5%; Score 89; DB 1; Length 212;  
 Best Local Similarity 24.2%; Pred. No. 0.75;  
 Matches 51; Conservative 37; Mismatches 95; Indels 28; Gaps 9;

QY 8 VELYDVLPSPSWLGFELCRYQNIWNIINLQRLPSLIITGIMKDSGNKPPCL--LPRKGLY 65  
 DB 16 VDFYDFLSPVSVLANHRLSKLAQDYGFPSIRYAYDLARVKTAGNVGPSNRDLIVKLDY 75  
 QY 66 MANDLKLRLHILPIHFDPKDFLSVMLEKSL--SAMRFLTA-VNLEHPMELEKASRELW 122  
 DB 76 LKVDLQWRABLYEPLVFPANYSRRMTGLYSGAMAQTGAYVNV-----VF 123  
 QY 123 MRVWSNEDITEPQSILAA--AEKAGMSAEQAQGLEKIATPKVKNQLEKTTAEACRYGA 180  
 DB 124 NAVW--GDGIAPDLESPLVSEKLGWDRSAFE---DFISSAATERYDEQTHAATERYK 178  
 QY 181 FGLPITVAHVGDQTHMFLGSDRMELLALLHLLG 211  
 DB 179 FGVP-TMFLGD---EMWGNDRFLMLENNAV 205

RESULT 7  
 SG2\_BOVIN STANDARD; PRT; 613 AA.  
 ID SG2\_BOVIN  
 AC P20616;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE SECRETOGNANIN II PRECURSOR (SGII) (CHROMOGNANIN C).  
 GN SCG2 OR CHGC.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-90264409; PubMed-2345170;  
 RA Fischer-Colbrie R., Gutierrez J., Hsu C.M., Tacangelo A., Eiden L.E.;  
 RT "Sequence analysis, tissue distribution and regulation by cell depolarization, and second messengers of bovine secretogranin II (chromogranin C) mRNA";  
 RT J. Biol. Chem. 265:9208-9213(1990).  
 CC -1- FUNCTION: SECRETOGNANIN II IS A NEUROENDOCRINE SECRETORY GRANULE PROTEIN, WHICH IS THE PRECURSOR FOR BIOLOGICALLY ACTIVE PEPTIDES.  
 CC -1- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY GRANULES.  
 CC -1- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.  
 CC -1- SIMILARITY: BELONGS TO THE CHROMOGNANIN / SECRETOGNANIN PROTEIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC EMBL; J05468; AAA30760.1;  
 DR PIR; A35296; A35296.  
 DR INTERPRO; IPR001990;  
 DR PFAM; PF01271; Granin; 1.  
 DR PROSITE; PS00422; GRANINS\_1; 1.  
 KW Sulfatation; Cleavage on pair of basic residues; Calcium-binding; Signal.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT PROPEP 28 30 POTENTIAL.  
 FT CHAIN 31 613 SECRETOGNANIN II.  
 FT PEPTIDE 181 213 UNKNOWN ACTIVITY PEPTIDE (PROBABLE).  
 FT MOD\_RES 150 150 SULFATATION (BY SIMILARITY).  
 SQ SEQUENCE 613 AA; 70356 MW; 5DC079F59D83516 CRC64;

Query Match 7.5%; Score 88.5; DB 1; Length 613;  
Best Local Similarity 23.6%; Pred. No. 3.1; Indels 59; Gaps 9;  
Matches 39; Conservative 26; Mismatches 41

QY 6 RTVELFYDLSVSWLGFELCRYQNIWNI-----NLQRLPSLTIGIMKSGNKPGLLP 60  
DB 333 RAIRLEKPLDP-----QSIYQLIEISRLQIIPPEDLIDMLK-TGEKP--VEP 377  
QY 61 RKGLYMANDKLLRHLLQIPIHPKFLSVLMLEKGSLSAMRFLAVNLEHPEMEKASRE 120  
DB 378 EQ-----ELEIPEV-PED-----ISEVDLDHDFLQFN----- 403

QY 121 LMRVWSRNEDETEPQISILAAAEKAGMSAEQAQGLL--EKIATPK 163  
DB 404 ---KMLSKNGYKPAKHGAVAEALPEGLSVEDILNLGMSAANPK 445

RESULT 8  
MIAA\_TREPA STANDARD; PRT; 316 AA.  
AC O83644;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE TRNA DELTA(2)-ISOPENTENYLPIROPHOSPHATE TRANSFERASE (EC 2.5.1.18) (IPP TRANSFERASE)  
GN MIAA OR TP0637  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN-NICHOLS;  
RX MEDLINE-98332770; PubMed-9665876;  
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;  
RA "Complete genome sequence of Treponema pallidum, the syphilis  
spirochete."  
RT Science 281:375-388(1998).  
RL -1- FUNCTION: CATALYZES THE FIRST STEP IN THE BIOSYNTHESIS OF  
CC 2-METHYLTHIO-N6-(DELTA(2)-ISOPENTENYL)-ADENOSINE (MS[2]I[6]A)  
CC ADJACENT TO THE ANTICODON OF SEVERAL TRNA SPECIES (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ISOPENTENYL DIPHOSPHATE + TRNA -  
CC PYROPHOSPHATE + TRNA CONTAINING 6-ISOPENTENYLADENOSINE.  
CC -1- SIMILARITY: BELONGS TO THE IPP TRANSFERASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: 249704; CAA89786.1; -  
CC SGD: S0004901; YMR288W.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 232 252 POTENTIAL.  
FT TRANSMEM 300 320 POTENTIAL.  
FT TRANSMEM 336 356 POTENTIAL.  
FT TRANSMEM 643 663 POTENTIAL.  
FT TRANSMEM 748 768 POTENTIAL.  
FT TRANSMEM 782 802 POTENTIAL.  
FT TRANSMEM 877 897 POTENTIAL.  
FT TRANSMEM 905 925 POTENTIAL.  
FT SEQUENCE 971 AA; 110027 MW; 27D26E4252A788E2 CRC64;

Query Match 7.1%; Score 84.5; DB 1; Length 971;  
Best Local Similarity 21.7%; Pred. No. 12;  
Matches 36; Conservative 31; Mismatches 64; Indels 35; Gaps 7;  
QY 5 PRVVELFYDLSVSWLGFELCRYQNIWNIQLRPSLTIGIMKSGNKPGLLP-KG 63  
DB 387 PYGIEVENVLEP-----LWKGIRSHRGVLSFLSKAVGSMPLMDPEYAG 432

QY 64 LYMANDKLLRHLLQIPIHPKFLSVLMLEKGSLSAMRFLAVNLEHPEML-EKASRELW 122  
DB 433 YTTTEAMRIIRREFDSPDEMKTKTLLVLOK--CSAVESIT-----PKFLREIEAIEFF 484

QY 123 MRVWSRNEDETEP-----QSIILAAAEKAGMSAEQAQGLLEKATP 162  
DB 485 QKFWVRVALDRPLNKVVVYTTVTLAKKLGSY-----TIDKLLTP 525

RESULT 10

Query Match 7.1%; Score 84.5; DB 1; Length 316;  
Best Local Similarity 23.3%; Pred. No. 3;  
Matches 41; Conservative 30; Mismatches 60; Indels 45; Gaps 8;  
QY 34 ATP (POTENTIAL).  
KW Transferase; Nucleotidyltransferase; trna processing; ATP-binding.  
FT NP\_BIND 27 34  
SQ SEQUENCE 316 AA; 36174 MW; E4BE144E56DE5208 CRC64;







FT MOD\_RES 153 153 SULFATATION (BY SIMILARITY).  
SQ SEQUENCE 619 AA; 71031 MW; 27CB7584F25A38D1 CRC64;

Query Match 6.9%; Score 81.5; DB 1; Length 619;  
Best Local Similarity 24.8%; Pred. No. 13;  
Matches 41; Conservative 16; Mismatches 57; Indels 51; Gaps 8;  
QY 13 DVLSPYSWLGFEILCRYQNIWININLQRLPSLITGIMKDSGNKPPGLPRKGLYMANDUKL 72  
Db 105 DVLSEDEM-----RIILEALRQAEENPPSALKENKPYALNLEK- 143  
QY 73 LRHHQLPIHPKDFSLVMEKGSLSAMRFLTAVNLEHPMELEKASRELWVRVSRNEDI 132  
Db 144 -----NFPVDPDDYETQWPERKLKHMRF-----PLMYEENSRE---NPKRINEI 187  
QY 133 TE-----QSIILAAEAKGMSAEQAQGLLEKIATPKVKNQKETE 173  
Db 188 VEEQYTFQS-LATLESVFEQ-----LGLKTGP--SNQKRERVD 222

RESULT 13  
DPOL\_ADEL2 STANDARD; PRT; 1061 AA.  
AC P06538;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DNA POLYMERASE (EC 2.7.7.7).  
GN POL.  
OS Human adenovirus type 12.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-94076430; PubMed-8254750;  
RA Sprengel J., Schmitz B., Heuss-Neitzel D., Zock C., Doerfler W.;  
RT "Nucleotide sequence of human adenovirus type 12 DNA: comparative  
functional analysis".  
RL J. Virol. 68:379-389(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-87106854; PubMed-3803925;  
RA Shu L., Hong J.S., Wei Y.-F., Engler J.A.;  
RT "Nucleotide sequence of the genes encoded in early region 2b of human  
adenovirus type 12".  
RL Gene 46:187-195(1986).  
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =  
N PYROPHOSPHATE + DNA(N).  
CC -!- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.  
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X73487; CAA51882.1;  
DR EMBL; M14785; AAA42478.1; ALT\_INIT.  
DR PIR; A25770; DJAD12.  
DR PIR; S33933; S33933.  
DR INTERPRO; IPRO02064;  
DR PRINTS; PR00106; DNAPOLB.  
DR PROSITE; PS00116; DNA\_POLYMERASE\_B; 1.  
KW Transferase; DNA-directed DNA polymerase; DNA replication;  
KW DNA-binding.

Query Match 32 R -> S (IN REF. 2).  
Best Local Similarity 162 V -> L (IN REF. 2).  
Matches 162 182 LQ -> YN (IN REF. 2).  
FT CONFLICT 181  
FT CONFLICT 461  
FT CONFLICT 461  
FT CONFLICT 575  
FT CONFLICT 575

FT CONFLICT 892 892 S -> T (IN REF. 2).  
FT CONFLICT 1030 1030 K -> M (IN REF. 2).  
SQ SEQUENCE 1061 AA; 121727 MW; 33FBA89C33065C08 CRC64;  
Query Match 6.9%; Score 81.5; DB 1; Length 1061;  
Best Local Similarity 21.8%; Pred. No. 25;  
Matches 50; Conservative 35; Mismatches 77; Indels 67; Gaps 12;  
QY 1 MGPLPRVVELF--YDVLSPYSWLGFEILCRYQNIWININLQRLPSLITGIMKDSGNKPPGL 58  
Db 133 IGSHPRPERLFIYDV-ETIWMG-----AFGKOLVFFML--VMKLSG----- 172  
QY 59 LPRKGLYMANDKLLRHLHQLPI-----HFPKX-----FLSVMEKGSLSAMRFLTAVNLEH 110  
Db 173 -----DDNLVKHALQALALELWQWDEKDSITFYCLTPEKMKV-GQOFRTYRN--- 218  
QY 111 PEMLEKASRELWVRVSRNEDITEPOSILAAEAKGMSAEQAQGLLEKIATPKVKNOLKE 170  
Db 219 -RLQTSLATDIWMTFLQKNPHLSQ-----WQAEENGLVALEDLSYEDLKRPAIKGEPR- 271  
QY 171 TTEACRYGAFGLPITVAHVDGQTHMLFGSDRMELLALLHLLGKWMGP 219  
Db 272 -----FVELYIVG--HNINGFDEIVLAAQVINRLDVP 304

RESULT 14  
YG73\_SYNY3 STANDARD; PRT; 274 AA.  
AC P74261;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE HYPOTHETICAL TRNA/RRNA METHYLTRANSFERASE SLR1673 (EC 2.1.1.-).  
GN SLR1673.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-97061201; PubMed-8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
MYAJIMA N., Hirose M., Sugiyama M., Sasamoto S., Kimura T.,  
Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
Shimpo S., Takeluchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
Synechocystis sp. strain PCC6803. II. Sequence determination of the  
entire genome and assignment of potential protein-coding regions".  
RL DNA Res. 3:109-136(1996).  
CC -!- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRMH FAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; D90913; BAA18355.1;  
DR INTERPRO; IPR001537;  
DR PFAM; PF00588; Spou\_methylase; 1.  
KW Hypothetical protein; Transferase; Methyltransferase.  
SQ SEQUENCE 274 AA; 29701 MW; A4176C1061CAA88 CRC64;

Query Match 6.8%; Score 81; DB 1; Length 274;  
Best Local Similarity 21.1%; Pred. No. 5.1;  
Matches 35; Conservative 32; Mismatches 61; Indels 38; Gaps 6;  
QY 7 TVELFYDVLSPYSWLGFEILCRYQNIWININLQRLPSLITGIMKDSGNKPPGLPRKGLYM 66  
Db 107 TMDQFWRSPPPHARLGL-VLERLQDPGNLGTILRTAAATGV-----EGIWL 151

Db 262 AACAIISQFEQHIRAVAGLPLGNTDRHSDCVMENLIGDDIEKVPAILCEKNAVLHLYGKK 320  
Search completed: February 15, 2001, 15:53:24  
Job time: 156 sec

QY 67 ANDL-----KLRHHIQIPIHPK---DFLSVMLEKGSLSAMRFLTAVNLEHPMLEKA 117  
Db 152 TADCVDPSPKVLRSAGSSLLPQQLQSLPPLLEKFTQGLQIATV-----PQA 203  
QY 118 SRELWVRWNRNEDITEPOSILAAEKAGMSAEQAQGLLEKIATPK 163  
Db 204 TQILW-----EIDFQRTIVIFGSEGGQLSAPVLELTHQVAIPQ 243

RESULT 15  
PURK\_BRUME STANDARD; PRT; 339 AA.  
AC P52559;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE ATPASE SUBUNIT (EC 4.1.1.21)  
DE (AIR CARBOXYLASE) (AIRC).  
GN PURK.  
OS Brucella melitensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Brucellaceae; Brucella.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-16M;  
RA Warren R., Hoover D., Hadfield T., Drazek S.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: POSSESSES AN ATPASE ACTIVITY THAT IS DEPENDENT ON THE  
CC PRESENCE OF AIR (AMINOIMIDAZOLE RIBONUCLEOTIDE). THE ASSOCIATION  
CC OF PURK AND PURE PRODUCES AN ENZYME COMPLEX CAPABLE OF CONVERTING  
CC AIR TO CAIR EFFICIENTLY UNDER PHYSIOLOGICAL CONDITION  
CC (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 1-(5-PHOSPHORIBOSYL)-5-AMINO-4-IMIDAZOLE-  
CC CARBOXYLATE - 1-(5-PHOSPHORIBOSYL)-5-AMINOIMIDAZOLE + CO(2).  
CC -1- PATHWAY: SIXTH STEP IN DE NOVO PURINE BIOSYNTHESIS.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE PURK / PURT FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U10241; AAA57003.1;  
KW Purine biosynthesis; Lyase; Decarboxylase.  
SQ SEQUENCE 339 AA; 37166 MW; 951B0DC5E637324E CRC64;

Query Match 6.8%; Score 80.5; DB 1; Length 339;  
Best Local Similarity 21.7%; Pred. No. 7.3;  
Matches 65; Conservative 25; Mismatches 96; Indels 113; Gaps 14;  
QY 3 PLPRIVELYDYVLS---PYSWLGFEILCRYQNIWNINQLRPSLITGIMKDSGNKPPGLL 59  
Db 47 PLPTSPMNMKMQSQAPPTSWLKRRLFCPPPHWK-----SLTASOKSFSFKAALK 99  
QY 60 PRK-GLYMANDKLRLHRLHLPKIPKDFLSVMLEKGS-----SAMRFLTAVNLEH 110  
Db 100 PRPGGSWMTFKRSSPSALGRANH-PQDRRLGYDGKGVRLASLDETQACNAFAIN-KA 157  
QY 111 PEMLE---KASRELWVRWNRNEDITEPOSILAAEKAG-----TPKVNOLKETTE 146  
Db 158 PAILEGFVEFEV-----SVIAARDRSGNVAIFDLAENVHVKDGLATST 202  
QY 147 -----MSAEQAQGLLEKIA-----TPKVNOLKETTE 173  
Db 203 VPAAISVQTAEAARAEKLLHALDYGVGLGFEFFVLKDTLLANEFAPRVHNS-GHWTE 261  
QY 174 AACRYGAF-----GLPI--TVAHVDGQTHMLFGSD-----RMELLAHLLEGK 213

